Midline Report for the Mixed-Methods Cluster-Randomized Controlled Trial of Impact Network's eSchool 360 Model in Rural Zambia

MARCH 2020

Thomas de Hoop | Hannah Ring | Garima Siwach | Paula Dias |
Gelson Tembo (Palm Associates) | Victoria Rothbard | Anaïs Toungui

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Midline Report for the Mixed-Methods Cluster-Randomized Controlled Trial of Impact Network's eSchool 360 Model in Rural Zambia

MARCH 2020

Thomas de Hoop | Hannah Ring | Garima Siwach | Paula Dias | Gelson Tembo (Palm Associates) | Victoria Rothbard | Anaïs Toungui

American Institutes for Research®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

8908_02/20

Contents

Tables

Page

Figures

Executive Summary

Low- and middle-income countries have made significant progress toward placing children into schools, but student learning and achievement are often dreadfully low (Berry, Barnett, & Hinton, 2015; Pritchett, 2013). Globally, approximately 387 million children of primary school age are not acquiring basic reading and mathematics skills, even though about two-thirds are in school and will complete primary school (United Nations Educational, Scientific, and Cultural Organisation [UNESCO], 2017).

Zambia faces many common educational challenges. Literacy rates among young Zambian adults ages 15–24 are 58.5% for females and 70.3% for males, despite an average of 7.7 years and 7.9 years of education, respectively (Central Statistical Office, Ministry of Health, & ICF International, 2014; UNICEF, 2015). Furthermore, public spending on education is low relative to other regional countries: Zambia spends about 1.3% of its gross domestic product (GDP) on education as compared with average expenditures of 5.6% of GDP in other southern and eastern African countries (UNICEF, 2015). Zambia also has a large, autonomous community schooling system that formed during the nation's transition from a socialist economy. The community schooling system has expanded during the past 20 years to increase education access in remote areas. Estimates suggest that the number of Zambian community schools has increased from 100 schools in 1996 to about 2,325 schools with 473,458 children in 2017 (Chimese, 2014; DeStefano, 2006; Ministry of General Education, Republic of Zambia [MoGE], 2016). However, many community schools in Zambia are staffed by untrained, underpaid teachers who teach a substandard curriculum and who lack management skills and school supplies. These schools need a cost-effective solution for delivering quality education in order to improve learning outcomes.

The eSchool 360 model implemented by Impact Network (IN) represents a promising approach to improving educational outcomes by incorporating three potentially high-impact components that could create important synergies: a standardized e-learning curriculum and pedagogy, ongoing teacher training and professional development, and community ownership. Each component could, on its own, have positive impacts on student outcomes by engaging the three main factors in the education system—students, teachers, and parents. Combining these components into a single program may be particularly effective by aligning all three factors toward improving the educational outcomes of students. Earlier research has suggested that these complementarities may be substantial, with even higher impacts from educational technology programs that include a strong focus on pedagogical practices (Muralidharan, Singh, & Ganimian, 2019).

The American Institutes for Research (AIR) has designed and is implementing a mixed-methods cluster-randomized controlled trial (cluster-RCT) to determine the effects of IN's eSchool 360 model. The study comprises three main evaluation components: an impact evaluation of the eSchool 360 model, an analysis of the cost-effectiveness of the eSchool 360 model, and a process evaluation of the expansion of the eSchool 360 model. To determine the impact of the program, we are using a cluster-RCT in which 63 eligible schools have been randomly assigned either to receive IN's eSchool 360 program (30 treatment schools) or not to receive the program (33 control schools). The population consists of children ages 6–9 who live in close proximity to the 63 selected community schools across three districts—Petauke, Sinda, and Katete—in Zambia's Eastern Province. The primary cognitive skills outcomes are aggregate scores on the early grade reading assessment (EGRA), early grade mathematics assessment (EGMA), the Zambian Achievement Test (ZAT), and Oral Vocabulary assessments.

This report presents the midline results of the cluster-RCT used to determine the impact of IN's eSchool 360 model. It examines the impact of the program on EGRA, EGMA, ZAT, and Oral Vocabulary assessments 14 months after the start of the program. In addition, it analyzes the impact of the program on intermediate outcomes, including school attendance and enrollment, perceptions of school and education quality, aspirations about child's education and marriage, child development, food security, and education expenditures. We present two main types of effects. The first, called the Intention-to-Treat (ITT) effect, considers the impact on all children in close proximity of the schools, regardless of whether they attended the school. The ITT identifies the impact of the *opportunity to attend* an IN school. The second, called the Treatment Effect on the Treated (TOT), estimates the effect for those children who attend an IN school when given that opportunity.

The estimated ITT effects suggest that the opportunity to attend an IN school positively affected children's learning outcomes across the board. On average, we found statistically significant effects of 0.40 standard deviations or 3.5 percentage points on EGRA scores, 0.22 standard deviations or 4.9 percentage points on EGMA scores, 0.16 standard deviations or 3.1 percentage points on ZAT scores, and 0.25 standard deviations or 6.0 percentage points on Oral Vocabulary scores 14 months after the start of the program. These effect sizes are equivalent to 1.88 additional years of education for reading and 1.03 additional years of education for mathematics if we extrapolate results of learning gains from Grade 1 to Grade 12 from a sample of representative low- and middle-income countries (Evans & Yuan, 2019). When compared directly against the one-year learning gains in the control group (who took the same tests and were measured on the same scale), the ITT estimates are equivalent to 1.2 additional years of education for reading and 0.37 additional years (or 4.5 additional months) of education for mathematics.

The estimated TOT effects suggest that the treatment effects on the treated for students whose parents chose for them to attend an IN school after getting the opportunity were substantially larger than the ITT effects. On average, participating in the eSchool 360 model resulted in improvements of 0.32 standard deviations or 6.3 percentage points in ZAT scores, 0.83 standard deviations or 7.2 percentage points in EGRA scores, 0.45 standard deviations or 10.1 percentage points in EGMA scores, and 0.52 standard deviations or 12.4 percentage points in Oral Vocabulary scores for students who reported that they attended IN schools three times in the last week. For students who reported that they had ever enrolled in IN schools (over the past year), we found TOT effects of 0.26 standard deviations or 5.2 percentage points in ZAT scores, 0.68 standard deviations or 5.8 percentage points in EGRA scores, 0.37 standard deviations or 8.2 percentage points in EGMA scores, and 0.42 standard deviations or 10.1 percentage points in Oral Vocabulary scores.

The results indicate that improved quality of education and increased school enrollment and attendance contributed to the positive effects on EGRA, EGMA, ZAT, and Oral Vocabulary scores. Residing in an IN catchment area increased the likelihood of school enrollment by 7.9 percentage points. We also found statistically significant effects of the eSchool 360 model on parents' satisfaction with education and school attendance. In addition, we found strong and statistically significant correlations between parents' satisfaction with education and school attendance with EGRA, EGMA, ZAT, as well as Oral Vocabulary scores.

We found less evidence for the contribution of other potential mechanisms to improvements in EGRA, EGMA, ZAT, and Oral Vocabulary scores caused by the eSchool 360 model. We did not find statistically significant effects of the eSchool 360 model on child development outcomes, including cognitive, social-emotional, and motor skills.

Qualitative respondents (students, teachers, parents, and program staff) also reported a high perceived quality of education at IN schools in interviews and focus group discussions. Their experiences suggest that IN schools provided a higher quality education than community schools and that teachers were present more regularly at IN schools than at government schools. They also attributed improvements in literacy outcomes to the IN schools. However, they did not report improvements in mathematics.

The qualitative research also showed evidence for a high level of fidelity of implementation, which likely contributed to the positive effects. The findings suggest that teachers used the participatory pedagogical approaches on which they were trained; that they adhered to the curriculum prescribed in the curriculum map; and that they used technology (tablets and projectors) as recommended.

Despite the positive program effects, the average reading and mathematics scores for children residing in IN catchment areas remained relatively low. They scored an average of only 10.9% correct on EGRA assessments and 24.3% correct on EGMA assessments, while children in the control group scored an average of 7.6% and 21.0% correct on EGRA and EGMA assessments, respectively.

The results of the longer-term impact and cost-effectiveness analysis will show evidence on whether and how the eSchool 360 model can be moved to scale. We will estimate cost data of the eSchool 360 model in the first half of 2020. In addition, we will estimate longer-term impacts following the endline survey three years after the start of the program. We will present these results in a future endline report.

Introduction

There have been dramatic increases in educational attainment worldwide during the past several decades, but the quality of education and overall student learning demand substantial continued improvement (Berry et al., 2015; Pritchett, 2013; World Bank, 2018). Globally, approximately 387 million children of primary school age are not acquiring basic reading and mathematics skills, even though about two-thirds are in school and will complete primary school (UNESCO, 2017). The current set of global development goals—the Sustainable Development Goals—shifts the focus from educational attainment to education quality, with the goal to "ensure inclusive and equitable **quality** education and promote lifelong learning opportunities for all" (United Nations, 2017, p. 24; emphasis added).

Zambia is emblematic of many low- and middle-income countries that face multiple educational challenges. First, overall education quality is low: literacy rates among young Zambian adults aged 15–24 are 58.5% for females and 70.3% for males, despite an average of 7.7 years and 7.9 years of education, respectively (Central Statistical Office, Ministry of Health, & ICF International, 2014; UNICEF, 2015). Second, Zambian public spending on education is low relative to other regional countries: Zambia spends about 1.3% of its GDP on education as compared with an average education expenditure of 5.6% of GDP in other southern and eastern African nations (UNICEF, 2015). Third, Zambia has a large, autonomous community schooling system that formed during the country's transition from a socialist economy. The system has expanded over the past 20 years to increase education access in remote areas. Estimates suggest that the number of community schools has increased from 100 schools in 1996 to about 2,325 schools with 473,458 children in 2017 (Chimese, 2014; DeStefano, 2006; MoGE, 2016). However, community schools are often staffed by untrained, underpaid teachers who teach a substandard curriculum and who lack management skills and school supplies. Improving education quality in community schools may be an effective entry point to improve educational outcomes for vulnerable children in remote areas.

This midline study focuses on the impact and fidelity of implementation of IN's eSchool 360 model 14 months after program start. The eSchool 360 model represents a potentially promising approach to delivering quality education and improving educational outcomes for students in community school in rural Zambia. It incorporates three potentially high-impact interventions that could offer important complementarities: a standardized e-learning curriculum and pedagogy, ongoing teacher training and professional development, and community ownership. The e-learning component provides electricity via solar power (supplied by IN), and projectors and tablets (supplied by IN's partner, *Mwabu*) for the community schools. The tablets are loaded with local-language materials that are structured around a curriculum approved by the Zambian government. IN supplements the technology by providing teacher coaching, observing teacher practices in the classroom, and creating community ownership. Locally hired teachers receive weekly training focused on using the technology and enhancing their pedagogical skills.

Combining e-learning, ongoing teacher training and professional development, and community ownership components into a single program may be particularly effective by aligning the incentives of students, teachers, and parents toward improving student educational outcomes. These components could each, on their own, have positive impacts on student outcomes. Earlier research has demonstrated that engaging all three factors—students, teachers, and parents—in the education system may be particularly effective because it could create important synergies. For example, an educational technology program in urban India that included a strong focus on pedagogical practices showed positive effects that were greater than the sum of those obtained from separate educational technology or pedagogical interventions (Muralidharan, Singh, & Ganimian, 2019). A cost-effectiveness analysis of the same program found that it also was cost-effective in improving learning outcomes. This is important considering that a recent review found that technology-based education programs may not be cost-effective, even if they produce large impacts on learning outcomes (Muralidharan et al., 2019; Piper, Zuilkowski, Kwayumba, and Strigel, 2016).

Joel impact school.

This study comprises three main evaluation parts: an impact evaluation of the eSchool 360 model, an analysis of the model's cost-effectiveness, and a process evaluation of the model's expansion. This report presents midline findings related to the impact and process evaluation. In addition, we have produced an inception report that details the design of the cost-effectiveness analysis (De Hoop et al., 2017) and a baseline report that summarizes the quantitative data collection at baseline (De Hoop, Brudevold-Newman, & Davis, 2018). We will collect the cost data in 2020 and will present an analysis of these data in the endline report.

Below we present research questions related to the impact evaluation, cost-effectiveness analysis, and process evaluation components of this study. Each part of the evaluation is designed to answer different but complementary questions.

Impact Evaluation

a. What is the effect of the eSchool 360 program on students' numeracy, preliteracy, and literacy skills?
b. Do students enrolled in the eSchool 360 program improve in numeracy and literacy skills?
c. Does the eSchool 360 program increase attendance and enrollment?
d. Does the eSchool 360 program lead to an improved perception of school and education quality among students, teachers, and parents?
e. Does the eSchool 360 program improve parental and children's aspirations?

Cost-Effectiveness

a. How cost-effective is the eSchool 360 program in improving literacy outcomes?
b. How cost-effective is the eSchool 360 program in improving mathematics outcomes?

Process Evaluation

a. Was the eSchool 360 program implemented as designed? If not, why was it not implemented as designed, what were the challenges to implementing it as designed, and how was it implemented?
b. How did the eSchool 360 program implementation vary by geography, culture, and time of year?
c. Did perceptions of the quality of teachers differ among students, parents, teacher supervisors, and teachers? If yes, how?

The remainder of this midline report is structured as follows: The report begins with a description of the background of this study and includes an overview of the existing literature on the impact of technology-based education programs. Next, this report presents a description of the eSchool 360 model followed by a description of its theory of change and the research designs for the impact and process evaluations. The report then offers a detailed overview of the quantitative and qualitative midline results and discusses conclusions.

Background and eSchool 360 Model

The current Zambian educational system faces many challenges in providing quality education and remains inaccessible to many Zambians under the age of 18 (i.e., 52.5% of the nation's population) (Central Statistics Office Zambia, 2013). An estimated 600,000 students attend nongovernmental, autonomous community schools that do not offer a full range of grades, are in poor condition, and are funded through minimal government funding of less than U.S. $91 per year (1,000 Zambian Kwacha) per school (DeStefano, 2006). The Zambian government has introduced evidence-backed interventions in government schools to improve school quality, such as the Teaching at the Right Level program. This intervention groups students according to learning level rather than by age or grade (Banerjee et al., 2016). However, the autonomous nature of community schools raises concerns related to whether they may be neglected from this quality push as well as whether a different set of interventions may be necessary to improve their quality.

Previous research demonstrates that multifaceted education programs such as IN's eSchool 360 model can improve learning outcomes. A comprehensive, systematic review on the impact of education programs in low- and middle-income countries concludes that successful education programs address constraints at multiple levels (Snilstveit et al., 2016), which can only be achieved by multifaceted education programs. Kremer, Brannen, and Glennerster (2013) also highlight the importance of adapting the curriculum to the child's level to improve the effectiveness of pedagogical interventions. Furthermore, Conn (2014) suggests that among interventions that include teacher training as a component, those with adaptive instruction had larger effect sizes than those without adaptive instruction. Finally, Muralidharan et al. (2019) find that in an RCT of a personalized, computer-aided afterschool instruction program in India, students in the treatment group made significant gains in mathematics and Hindi test scores. The authors concluded that the impact was due primarily to the computer-aided learning system's ability to target and adapt to wide variations in student learning levels.

In contrast, there is evidence that programs in low- and middle-income countries that focus on increasing educational inputs without addressing other learning constraints are not sufficient to improve learning outcomes (Schling & Winters, 2018; Snilstveit et al., 2016). An increased provision of traditional school resources such as textbooks or flipcharts had no impact on student attainment (Glewwe, 2002). Banerjee, Cole, Duflo, and Linden (2007) note that increasing inputs to schooling fails to have an impact on student attainment if what is being taught remains too difficult for students to learn. Similarly, numerous studies that have focused on computer-assisted learning programs did not find significant impacts when the program focused only on increased educational inputs (Snilstveit et al., 2016; Valk, Rashid, & Elder, 2010).

The eSchool 360 Program. IN developed the eSchool 360 model to deliver low-cost education to children in rural communities through a holistic solution. A previous evaluation of the eSchool 360 model showed that the cost of the program was $3 per month per student, which is 70% less than the Zambian government spends per student (Winters, Schling, & Winters, 2013). The core of the eSchool 360 model is e-learning technology whereby tablets and projectors, provided by IN's partner, *Mwabu*, are loaded with both lesson plans for teachers and with interactive lessons for students, approved by the Zambian government, and delivered in the local language. IN supplies electricity via solar power and supplements the technology by providing teacher training and professional development and creating community ownership. Locally hired teachers receive weekly training focused on using the technology and enhancing teachers' pedagogical skills. The approach represents a significant innovation not only because technology is used but also because it incorporates the practice of training local high school graduates to be teachers and provides them with systematic, ongoing support.

Encouraging results of a previous nonexperimental but longitudinal evaluation of the eSchool 360 model led AIR to fund the expansion of the eSchool 360 model to 30 additional community schools in rural Zambia (from an initial core of nine community schools, making for a total of 39 community schools receiving the program).[1] The earlier analysis indicated that improving mathematics outcomes by 1 percentage point cost IN schools 88% less than it cost government schools, suggesting that the model may be cost-effective in improving learning outcomes in Zambia (Schling & Winters, 2018). IN conducted the expansion in 2017 by implementing the eSchool 360 model in 30 community schools across three rural Zambian districts (Katete, Sinda, and Petauke) in areas with no running water and limited electricity. The cohort of students that was studied for our evaluation was admitted into the program in January 2018.

[1] AIR decided to fund the program's expansion to include five quasi-governmental schools that were not participants in the cluster-RCT. We will compare learning outcomes of the students in these five schools to the learning outcomes of students in five comparable governmental schools during the final data collection, which will be three years after the introduction of the eSchool 360 model in the quasi-governmental schools.

AIR and IN designed the expansion to include a rigorous mixed-methods cluster-RCT to determine the impact of the program on students' learning outcomes. To achieve this goal, AIR and IN closely consulted with Zambian government officials to obtain letters of approval for random assignment of the eSchool 360 program to 30 treatment schools and 33 control schools.[2] The random assignment of schools was conducted in May 2017. The Zambian Ministry of Education officials implemented the randomization, while AIR staff ensured the integrity of the process. We chose an unbalanced design with a smaller number of treatment schools because of limited resources to implement the eSchool 360 model.

Theory of Change

The program's theory of change suggests that the eSchool 360 program may lead to improvements in learning outcomes through various mechanisms (Figure 1). First, the teacher professional development component of the model may produce improvements in knowledge and practices of untrained teachers, which may result in improvements in the quality of education—for example, through the integration of activity-based learning methods and improvements in the curriculum. These improvements, in turn, may lead to improvements in preliteracy, early grade reading, and early grade mathematics outcomes. Second, the infrastructure improvements in the community school may lead to increased demand for education, which may result in increased education enrollment and attendance. The infrastructure improvements may also result in decreases in the age-at-enrollment of Zambian students. These improvements in school attendance and enrollment may then result in an increase in the time spent on education, which may lead to improvements in learning outcomes. Third, the incorporation of technology may enable untrained teachers to be more effective because pre-prepared lessons are sequenced and built into the tablets, which are regularly checked and monitored by operations management.

In addition to the improved learning outcomes, the program may result in improvements in the aspirations of students and parents. Improvements in the quality of education may increase expectations for students' futures. These increased expectations may, in turn, lead to higher

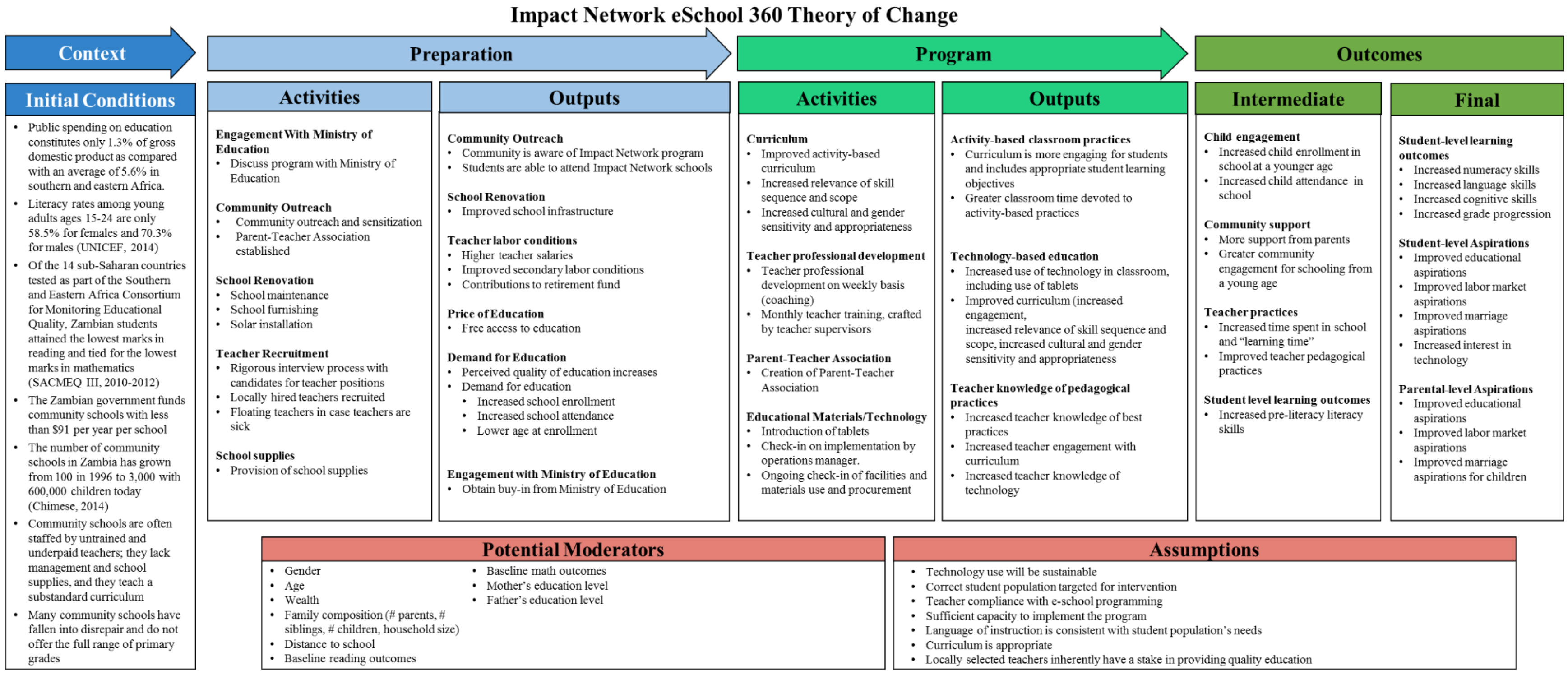
[2] An earlier version of this report included 34 control schools and 30 treated schools. However, discussions with program implementers from Impact Network showed that researchers had mistakenly included an excluded school to the treatment schools and a treatment school to the control group in the baseline analysis and the earlier report. Impact Network had initially identified 64 eligible schools (as opposed to 65 schools as reported in the inception and baseline report). However, further analysis suggested that the 64th school was too far away from the Impact Network offices to be considered for the program. This school had initially been assigned to the treatment group, but we replaced the school with a randomly selected school from the control group in the same district. This change was initially not taken into consideration in the analysis and baseline and midline data were collected in the catchment areas of this school as well. The current report presents results after correcting the sample composition (we removed the ineligible school from the analysis and switched the control group school to the treatment group for which it was earlier selected), and therefore includes slightly different sample sizes and impact estimates in comparison with the previous report.

aspirations in the domains of education, labor market, and family outcomes. For example, parents may increase their expectations of the likelihood that their children will be able to finish school. In addition, the improved quality of education may result in increased expectations of the returns of education, which may lead to higher expectations for labor market outcomes. Finally, increased aspirations in the education and labor market domains may result in increases in expectations relative to students' marriage prospects and increased age at marriage.

The validity of eSchool 360's theory of change depends on several assumptions. Perhaps most important, teachers need to comply with eSchool 360 programming. In addition, the community schools must have sufficient capacity to implement the model. Furthermore, locally selected teachers need to have appropriate incentives to provide quality education. The language of instruction also must be consistent with the needs of the student population.

The effects of the model may also vary with several individual-, household-, and community-level moderators. For example, the effects may vary by gender, age, and socioeconomic household-level characteristics. In addition, the model may be less effective in improving school attendance and enrollment for students who live farther away from the IN school. Finally, the model's impacts may be moderated by student baseline preliteracy, reading, and mathematics outcomes, as well as parental education levels. We test each of these potential heterogeneities in the impact evaluation.

Figure 1. Theory of Change



Study Design and Methods

AIR has designed and is implementing a mixed-methods cluster-RCT. The data collection began with quantitative baseline data, followed by quantitative and qualitative data collection 14 months after the start of the baseline data collection. We will collect endline data three years after program start. In addition to the quantitative and qualitative impact analyses, we will conduct a cost-effectiveness analysis. Specifically, we will assess the costs of the eSchool 360 model using the ingredients method. For this purpose, we will need to specify all ingredients that are necessary to replicate the model and then collect data on the unit costs of these ingredients (Dhaliwal, Duflo, Glennerster, & Tulloch, 2011). AIR will work with IN to gather information on resources used for the intervention to create an exhaustive list of resources with costs. Using this information, AIR will create a cost database. We will then estimate the costs of the intervention for the average beneficiary and divide these costs by the expected gain in outcomes derived from the impact analysis to serve as the cost-effectiveness measure of the intervention.

This report presents findings from the midline study, which comprised quantitative and qualitative data collection. The inception report includes a description of the methods we propose to use for the cost-effectiveness analyses (De Hoop et al., 2017).

Impact Evaluation Design

Sampling and Randomization

The cluster-RCT evaluation of the eSchool 360 model involved randomly assigning the program among schools that satisfied IN's geographic, infrastructure, and organizational structure eligibility criteria for the eSchool 360 expansion. The geographic criteria arose from IN's goal of introducing community schools across three districts—Petauke, Sinda, and Katete—in Zambia's Eastern Province. Of the schools in these districts, IN sought those with a dedicated physical structure, that were largely informal, and that had more community teachers than government teachers. Pairs of eligible schools that operated within three kilometers of one another were excluded to reduce bias from spillovers or contamination, which may occur if a child intended to be unaffected by the eSchool 360 model subsequently attends an IN school.

IN and AIR first consulted with local Zambian government officials to obtain a list of all community schools in the region. Of the 149 community schools that were identified, 63 met all eligibility criteria. IN staff then visited each of the 63 schools to obtain information on the structure of the school, the number of government and volunteer teachers, the state of the infrastructure, the grades served, and distance to other schools. Local representatives of the

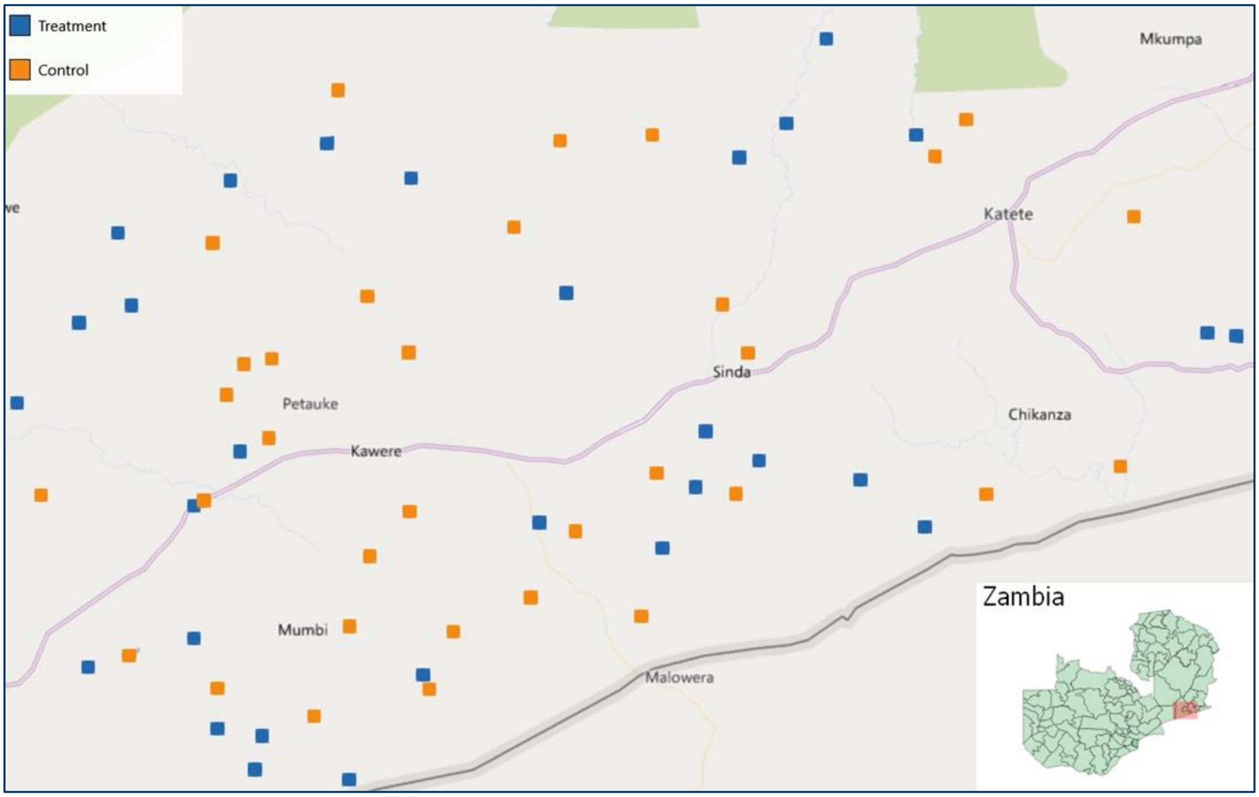
Ministry of Education implemented the randomization with oversight from AIR, assigning five schools in Katete, nine in Sinda, and 16 in Petauke to the treatment group (the group receiving the eSchool 360 model). The remaining 33 schools were assigned to the control group (the group not receiving the model).

This study estimates the impact of the eSchool 360 model on children who are eligible to enroll in first grade and who live near one of the 63 schools. We focused on these children because the introduction of the eSchool 360 model was designed to expand to an additional grade each year of the model's operation. In the first year, only the first-grade cohort received the full program package. This will expand to Grades 1 and 2 in the second year, and so on. The study will use a longitudinal panel design that follows each of the sampled children for three years, regardless of where and when they enroll in school, to estimate (1) Intention-to-Treat, or ITT effects of the model, which identifies the effect of the opportunity to enroll in an IN school; and (2) Treatment-Effects-on-the-Treated, or TOT effects, which identify the effect of attending an IN school on those who actively participate in the program (i.e., attend IN schools).

For the ITT analysis, we compared all eligible children who lived near the 30 IN schools with eligible children who lived near the 33 control schools, thus including all children regardless of whether they enrolled in a school. To identify the sample of children with the potential to be affected by the model and to then determine ITT effects, we conducted a census in the areas surrounding the sample schools to identify all households with children eligible to enroll in first grade in January 2018; that is, children ages 6 years or older in January 2018 who did not attend first grade or higher in the prior school year. We identified all households with children eligible to enroll in first grade within a diameter of 1.5 kilometers of the schools. We iteratively expanded the distance by 0.5 kilometers in communities with insufficient numbers of eligible children within the initial or subsequent sampling areas until we found sufficient eligible households. This procedure was implemented consistently across treatment and control school-catchment areas.[3]

[3] The initial distance stemmed from conversations with local experts who suggested that children eligible for first grade generally do not walk more than 1.5 kilometers to school. Since those conversations were conducted, we have identified a number of cases in which children walk longer than 3 hours to attend school.

Figure 2. Map of Sample Schools



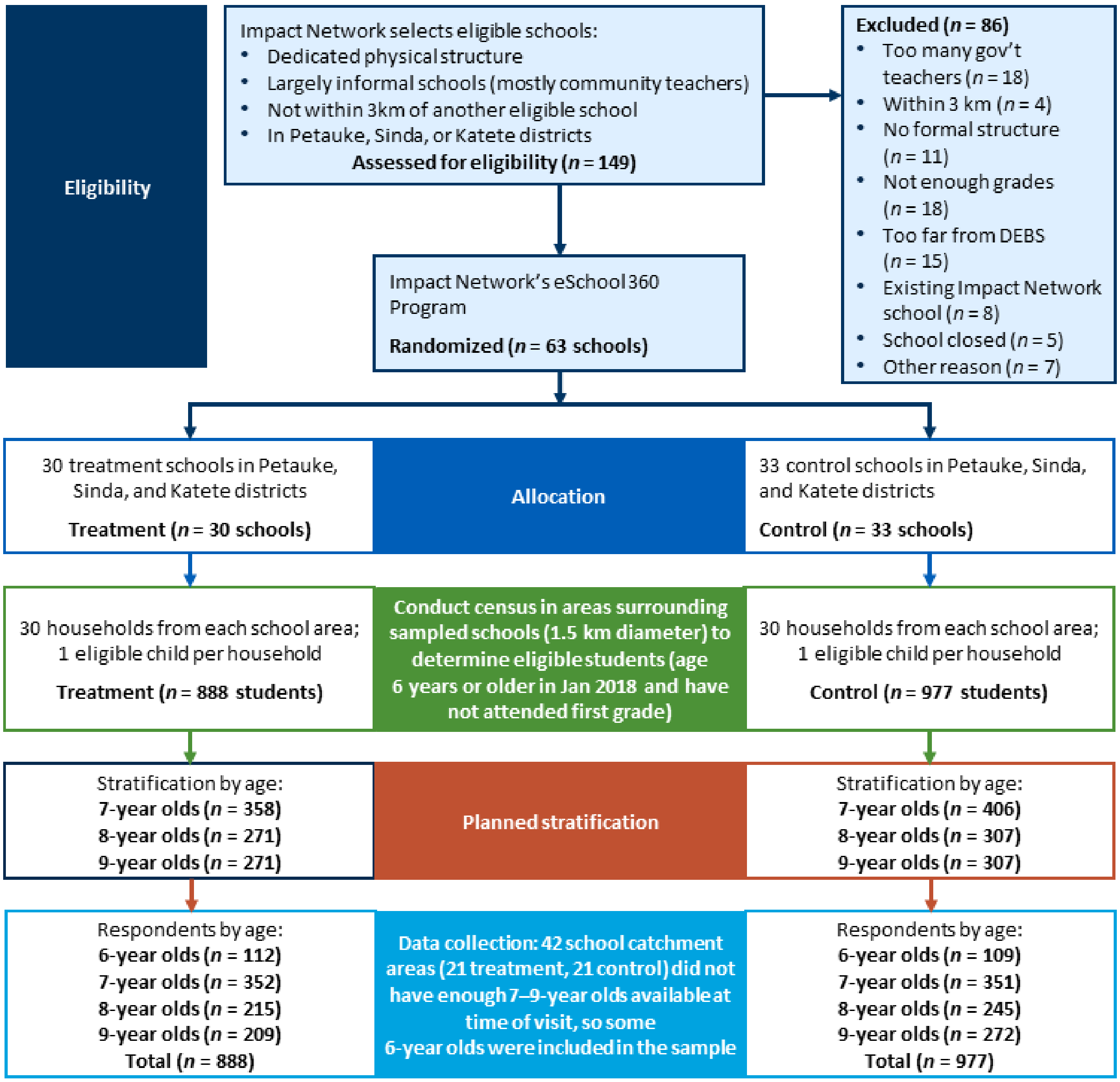
It is important to identify ITT effects because the announcement of the treatment schools and the construction of the additional infrastructure involved with the implementation of the eSchool 360 model are visible to the beneficiaries. The randomization of schools to treatment occurred nearly seven months before the first evaluation cohort was admitted to the sample schools in January 2018. This visibility may result in a different composition of students attending the treatment schools relative to the control schools—for example, by influencing school enrollment and attendance. Such changes in the composition could result in a bias in the impact estimator that compares students enrolled in the IN schools with students enrolled in other community schools because of differences in either observable (e.g., age, gender, parents' education level) or unobservable (e.g., motivation, noncognitive skills) characteristics.

To address concerns regarding these composition effects and to enable the estimation of ITT effects, we randomly sampled 30 households from the census-generated sample frame for each of the sample schools. For households with more than one eligible child, we selected the oldest child for inclusion in the sample. The introduction of IN schools may have induced some eligible children to enroll in school (who would not have enrolled in school absent IN). Comparing the outcomes (e.g., test scores) of all eligible children allows for capturing any effects on inducing children to attend school.

To increase statistical power, we oversampled 8-year-olds and 9-year-olds from our sample based on descriptive statistics about the age distribution from the Eastern Province (Ministry of General Education, 2014). These descriptive statistics indicated that 8-year-olds and 9-year-olds are more likely than 7-year-olds and 6-year-olds to be enrolled in first grade (Ministry of General Education, 2014). For our sampling strategy, we began by randomly selecting households with 7-year-olds, 8-year-olds, and 9-year-olds within a distance of 1.5 kilometers from each of the schools. However, we had to slightly adjust the sampling strategy because 42 catchment areas did not have sufficient numbers of 7- to 9-year-old children.[4] The sample was adjusted by randomly sampling additional 6-year-old children, followed by sequentially expanding the radius of the circle by an additional 0.5 kilometers until a sufficient sample was identified. Figure 3 presents the consort flow chart that describes the sampling strategy, starting with the eligibility criteria, followed by the random assignment, the census, the planned stratification by age, and the actual stratification following the practical decisions that we made during data collection.

[4] A "catchment area" is defined as the geographic area served by a particular school.

Figure 3. Consort Flow Chart



Quantitative Data Collection

AIR is partnering with Palm Associates, a Zambian research organization that specializes in data collection and social science research, to collect data for the evaluation of IN's eSchool 360 model. AIR worked closely with Palm to train enumerators before the baseline and midline data collection and then followed enumerators into the field to observe data collection. Palm works with Zambian enumerators who speak the local language of the areas included in the study and who are familiar with the assessments that AIR will implement.

Palm used tablets to conduct the baseline data collection, thus improving the quality of data collection, minimizing the need to clean data, and eliminating data entry. We collected the data on tablets running SurveyCTO software, which minimizes errors in the field because skipping patterns can be automated and built-in checks ensure the quality of data. In addition, the tablets allow for the collection of GPS data, which enables the tracking of households from baseline to midline data collection in the field.

The primary achievement indicators for numeracy, preliteracy, and literacy come from the EGRA, EGMA, and ZAT, which have already been adapted and used in the same region as this study. In addition, we collected Oral Vocabulary assessment data on the students. These instruments have been translated and validated in the context of Zambia. We collected assessment data primarily in the Nyanja and Chewa dialects because these are the languages of instruction in Grades 1–3 in Zambia's Eastern Province. Further, we collected enrollment and attendance data during the midline household survey as well as outcome indicators on parental and community perceptions of schools, teachers, and their children's education. Finally, we collected outcome data on parents' and children's educational, labor market, and marriage aspirations. In addition to these outcome indicators, the survey incorporated other moderating variables at different levels, including (1) student-level—gender and age; and (2) household-level—distance from school, poverty level, parents' education level, and household size. During the endline data collection, we also will collect data at the school level, including data on school size, number of teachers, number of teaching years, teacher age, and average class size.

The baseline report included a detailed overview of the construction of outcome variables (De Hoop et al., 2018).

Cluster-RCT Analysis

The main impact analysis used an ITT analysis to measure the impact of living within 1.5 kilometers (or within three kilometers for a small minority of the children) of an IN school on student academic outcomes, school attendance, and perceptions of the quality of education, regardless of whether the student chose to attend the IN school. We used an ANCOVA model to estimate the ITT effect 14 months after the start of the program. The ANCOVA approach uses a regression specification that includes the baseline measures of outcome variables as an additional explanatory variable.

The proposed evaluation design, with random assignment of schools to treatment, provides an unconfounded measure of the direct effect of the IN intervention on student outcomes. The probability of assignment to treatment is orthogonal to individual characteristics after controlling for stratum fixed effects. Thus, the direct effect of treatment (residing in an IN school catchment area) on outcome Y_i can be estimated using the regression specification:

$$Y_{i2019} = \alpha + \beta\, IN_i + \delta S_i + \sigma Y_{i2018} + \mu\, C_i + \epsilon_i$$

Here, IN_i is an indicator variable equal to 1 if individual *i* resides in the catchment area of an IN school and equal to 0 otherwise, S_i is a vector of dummy variables for the district strata, Y_{i2018} is a baseline value of the outcome of interest, C_i is a vector of other control variables, and ϵ_i is a conditionally mean-zero error term. Since treatment was randomized within strata, the inclusion of the Y_{i2018}, S_i, and C_i variables should increase efficiency but not impact the estimated value of β. If, by chance, there is some imbalance in the baseline covariates between the treatment and control groups, the regression model adjusts the outcome values to account for these differences, which leads to a more precise estimate of β. We used cluster-robust standard errors clustered at the school level to account for potential correlation in outcomes within a school catchment area. To bring the sample to being representative of the population in terms of likelihood of enrolling in school by age, we applied post-stratification weights to our analyses. We report all our impact estimates in two units – the original score unit measured in terms of percentage-points (as a score out of 100), and standardized scores where each student's midline score is standardized to the midline mean and standard deviation of the control group.[5]

To address the potential inflation of Type I error and statistical significance owing to multiple comparisons, we applied corrections for multiple comparisons to multiple outcome measures within the same outcome domain using the Benjamini-Hochberg procedure, as recommended by the What Works Clearinghouse and employed by Banerjee et al. (2015). The outcome measures for the impact analyses are organized into two domains: primary EGRA outcomes and primary EGMA outcomes. These outcomes are single measures and are not corrected for multiple comparisons. The EGRA and EGMA subtask outcomes are related within the domains, and the results are corrected for multiple comparisons. We also applied corrections for multiple comparisons in the estimation of heterogeneous effects. Finally, we conducted a robustness check by applying randomization inference, which examines whether the results are robust to any variations in the data that may have arisen from the randomization by calculating the proportion of alternative realizations of the regression coefficient that are further away from zero than the estimated regression coefficient (Heß, 2017; Young, 2019).

[5] We chose to standardize against the midline score of the control group because we consider the midline score of the control group more relevant and more representative for first-grade children in community schools in Zambia than the baseline score for the full sample. Some other impact evaluations of education programs in international development chose to standardize against the baseline sample mean and standard deviation (see for example, Muralidharan et al., 2019). However, in our study, none of the children attend school at baseline, leading to little variation across the sample. Using the baseline sample mean and standard deviation for standardization would thus result in inflated effect sizes when we would communicate effect sizes in standardized mean differences.

We estimated treatment effects on the treated, which reflect the effects of attending an IN school, by using the program assignment as an instrumental variable for self-reported school attendance in IN schools. The TOT is the impact of the intervention on those children who participated in the eSchool 360 model. Within this group we distinguish between children who actively participated in the program (children who self-reported attending IN schools at least three times in the week before the survey), and children who had ever enrolled in IN schools in the last year (children who may have participated in the eSchool 360 model but likely less so than the subgroup of children who attended IN schools at least three times in the week before the survey).

The theory of change identified a number of potential moderators to the impact of the IN model. To account for these moderators, we tested for whether the program had a differential impact by age; gender; region; socioeconomic status; mother's education level; and baseline EGRA, EGMA, and ZAT attainment. For noncategorical moderators, we constructed a binary variable equal to 1.0 for individuals with a value greater than the median and equal to 0.0 for those with a value less than the median. We then interacted the binary variable with treatment to test for whether the treatment was statistically different for groups with high and low levels. For socioeconomic status, we constructed an asset index using the calculated values from the first principal component of a list of assets as recommended by Filmer and Pritchett (2001).

Estimating Equivalent Years of Schooling

Finally, we transformed the effects on EGRA and EGMA outcomes to equivalent years of schooling using two different methods. For the first method, we used estimates from Evans and Yuan (2019), who show that students learn between 0.15 and 0.21 standard deviation of literacy ability over the course of a business-as-usual school year in a sample of representative low- and middle-income countries. Specifically, we use the estimates from their regression model for the pooled sample of Bolivia, Colombia, Ghana, Kenya, and Vietnam. The model explains reading proficiency with the number of years of schooling and a number of control variables and estimates the learning gain from an additional year of schooling based on the learning gain from Grade 6, usually the last year of primary school. We used the regression model that does not include country fixed effects, which suggests that 4.7 years of schooling results in a one standard deviation improvement in reading proficiency.[6] For the second approach we used estimates of the learning gains in the original test scale for the control group since the baseline survey. These learning gains are a weighted average of learning gains for children who do and who do not attend primary school in the year before the survey. This

[6] Evans and Yuan (2019) standardize scores against the mean and standard deviation of the population scores, and therefore the comparison between our effect sizes and results from their study may not be perfectly reconcilable. Nevertheless, we report the results to enable a comparison with other education programs in international development settings.

comparison may be more valid for interpreting learning gains against a counterfactual setting, since all children in the control group are tested on the same assessment and are scored on the exact same scale as the treatment group.

Process Evaluation Study Design

AIR is also implementing a process evaluation to assess the fidelity of implementation of the eSchool 360 program. The process evaluation complements the impact evaluation by enabling us to explore informal patterns and unanticipated interactions in addition to formal activities and anticipated outcomes, while also giving us flexibility to explore unforeseen areas of interest. Incorporating a qualitative process evaluation into a larger impact evaluation allows us to investigate why we do or do not see impacts and to explore nuances and variations in our sample, particularly with regard to sensitive or complex topics. Further, this component of the study provides rich contextual information to complement the numerical findings of the impact evaluation while also allowing us to triangulate and explain quantitative findings.

Qualitative Data Collection

We used three primary approaches to collect qualitative data for this evaluation: key informant interviews (KIIs) with teachers, teacher supervisors, and IN program staff; focus group discussions (FGDs) with students and parents; and classroom observations.

KIIs—KIIs with carefully selected teachers, teacher supervisors, and IN staff are an important element of our evaluation design because they shed light on perceptions of the quality of the eSchool 360 model both in terms of how the technology and teacher trainings have affected classroom instruction and how the program engages with school communities. Throughout KIIs, respondents discussed the perceived impact of the program, and we explored respondent beliefs about whether the program is being implemented as planned and students are being instructed as intended.

FGDs—FGDs provide a context in which participants feel comfortable and empowered to discuss the evaluation topics with their peers and the carefully trained facilitators. We created a social dynamic that encourages participants to reflect upon their opinions and experiences and express them verbally. FGDs provided an opportunity to gain a deeper understanding of the factors enabling or inhibiting implementation of the eSchool 360 curriculum in the classroom. From FGDs, we investigated experiences that parents and PTA members had when communicating with IN staff, generally engaging with the program, and using the school in the evenings and on weekends. With students, we discussed the techniques that teachers used in the classroom as well as the friendliness of the school environment.

Classroom Observations—We conducted formal classroom observations at one treatment school in each of the three districts. Classroom observations enabled us to examine the extent to which teachers at IN schools were effectively using e-learning technology and applying pedagogical techniques gained during eSchool 360 teacher trainings. Observations also noted the eSchool 360 environment, particularly with respect to its child-friendliness. We conducted classroom observations in one randomly selected academic class at each of the three selected schools using a piloted and validated instrument.

KIIs, FGDs, and classroom observations took place in one treatment school in each of the three districts of Katete, Petauke, and Sinda. In each district, the research team worked with IN staff to select one school to visit. The selection was made based on a list of observable characteristics including school size, distance from district center, and the distribution of high/low performing schools based on student learning outcomes from prior years. Table 1 provides a summary of the qualitative data collection approaches and respondents.

Table 1. Summary of Qualitative Methods for eSchool 360 Process Evaluation

Method	Respondents	Number
FGD	Students	3
FGD	Parents/PTA members	3
KII	Teachers	3
KII	Teacher supervisors	3
KII	IN staff	8
Classroom observation	N/A	3
Total		23

AIR partnered with Palm Associates to conduct qualitative data collection. Two experienced enumerators were recruited by Palm Associates to carry out interviews, focus groups, and classroom observations. Researchers spoke the Nyanja and Nsenga dialects, and qualitative instruments were translated into these languages during a three-day training. The researchers digitally recorded all KIIs and FGDs on portable digital recorders. They downloaded recordings to field laptops each day, renamed according to an anonymized code system held in an encrypted spreadsheet and copied to external media for backup. At the end of each day, the field researchers transcribed the handwritten field recording sheets to word processed documents, translating the material as necessary.

Researchers used audio recordings to supplement and validate the written transcriptions and translations. They assigned all transcriptions new names according to the code system in order to ensure data and informant confidentiality. The data collection supervisor ensured that the work was carried out in accordance with research plans, on schedule, ethically, safely, and completely. The supervisor also reviewed transcribed interviews to ensure high standards of data quality.

Qualitative Data Analysis

All data from KIIs and FGDs were coded and analyzed using the NVivo qualitative software program. Our team created a preliminary coding outline and structure based on the research questions, interview protocols, and memos of ideas that emerged during data collection. This coding outline served as a tool to organize and subsequently analyze information gathered in the interviews and focus groups. We created a list of definitions for the codes to ensure that coders categorized data using the same standards. After inputting the raw data into NVivo, coders selected a sample of interviews to double-code in order to ensure interrater reliability. Using these coded data, the qualitative team applied grounded theory to identify themes, categories, and theories that emerged from the data and that confirmed or refuted the researchers' initial impressions. In other words, rather than basing the analysis on a hypothesis, the researchers created concepts and categories based on the data, refining the concepts throughout the process to ultimately inform the overall findings. During this process of data reduction, researchers characterized the prevalence of responses, examined differences among groups, and identified key findings and themes related to the research questions.

Results

Baseline Equivalence

We did not find statistically significant differences in the large majority of background characteristics across households residing in IN catchment areas and control households. Child and household-level characteristics, including district of residence, household perceptions of poverty, and household distance from school were similar across treatment and control households. Overall, almost half of the sample was age 8 years or older, and slightly more than half of the sample was male. Most of the sample resided in Petauke District (compared with Katete or Sinda) and considered itself to be "very poor." More than half of the sample reported that the child caregiver had attended school, and the average distance from school was approximately 0.77 kilometers. We report descriptive statistics on sample background characteristics by treatment and control areas in Table 2.

Although we found some statistically significant differences between treatment and control households, only the difference in household distance from school was larger than 0.25 standard deviations. A study by Imbens (2015) indicates that statistically significant differences that are smaller than 0.25 standard deviations are unlikely to cause bias. However, we did observe some small but statistically significant differences when comparing sample means. For example, children in treatment households were less likely to be age 8 years or older (48%) compared with control households (53%). We also found that children in control group scored higher on baseline EGMA scores than children in the treatment group. All regressions estimating the ITT and TOT effects control for variables that showed some imbalance at baseline to generate more precise estimates. Specifically, we control for household distance from school, baseline EGMA score, and child age at baseline.

Table 2. Baseline Equivalence

Variables	Control		Treatment		Difference Test			Std. Mean Difference
	Mean	*N*1	Mean	*N*2	Diff	SE	*p*-Value	
Background characteristics								
Index child was female	0.48	888	0.44	812	-0.04	0.02	0.08	-0.08
Index child was 8 years old or older at baseline	0.53	888	0.48	812	-0.06	0.03	0.04	-0.11
Caregiver had attended school	0.60	880	0.66	808	0.06	0.03	0.06	0.13
Resided in Katete District	0.15	888	0.17	812	0.02	0.09	0.87	0.04
Resided in Petauke District	0.60	888	0.54	812	-0.06	0.13	0.66	-0.11
Resided in Sinda District	0.25	888	0.29	812	0.04	0.11	0.72	0.09
Household considered itself to be nonpoor	0.02	888	0.01	812	-0.00	0.01	0.61	-0.03
Household considered itself to be moderately poor	0.49	888	0.51	812	0.02	0.03	0.44	0.05
Household considered itself to be very poor	0.49	888	0.47	812	-0.02	0.03	0.53	-0.04

Variables	Control		Treatment		Difference Test			Std. Mean Difference
	Mean	*N*1	Mean	*N*2	Diff	SE	*p*-Value	
Household distance from school (km)	0.68	888	0.88	812	0.20	0.10	0.06	0.27
Primary outcomes								
Zambian Achievement Test (% correct)	0.45	888	0.44	812	-0.02	0.02	0.48	-0.07
Early Grade Reading Assessment (% correct)	0.05	888	0.05	812	0.00	0.00	0.99	0.00
Early Grade Mathematics Assessment (% correct)	0.08	888	0.06	812	-0.02	0.01	0.03	-0.18
Oral vocabulary (% correct)	0.62	888	0.59	812	-0.03	0.02	0.20	-0.11

Note. Standard errors (SE) clustered at the school level.

Attrition and Program Take-Up

Midline Attrition—During the midline survey, several logistical challenges delayed data collection activities. First, we experienced limited access to areas due to poor road conditions and heavy rains. Second, poor network connectivity prevented continuous communication with teams working in remote areas. Third, it was challenging to track households and children who migrated after the baseline. We tracked and administered surveys with 91.61% (*N* = 1,700) of the 1,865 children in the baseline sample. Table 3 below presents a summary of the response status by district. The table classifies each household into one of five categories: (1) Complete interview, (2) Partially complete interview, (3) Noncontact, (4) Refusal, and (5) Other. We provide a detailed definition for each category under Table 3.

Table 3. Summary of Response Status by District

Interview Response Status	Katete		Petauke		Sinda		Total Number of Interviews
	N	%	*N*	%	*N*	%	
Complete interview	273	91.61%	967	94%	460	85.66%	1,700

Interview Response Status	Katete		Petauke		Sinda		Total Number of Interviews
	N	%	*N*	%	*N*	%	
Partially complete interview	9	3.02%	11	1%	10	1.86%	30
Noncontact	8	2.68%	16	2%	37	6.89%	61
Refusal	0	0%	1	0%	0	0%	1
Other	8	2.68%	35	3%	30	5.59%	73
TOTAL	298		1,030		537		1,865

Note. "Complete interview" means that both the primary caregiver and the index-child were interviewed by an enumerator. "Partially complete interview" means that the primary caregiver was interviewed but not the index-child. "Noncontact" means that both the primary caregiver and the index-child were not interviewed, even though the household from baseline was located. "Refusal" means that the household refused to participate in the survey. Most of the interviews marked as "Other" either relocated or could not be identified using the information collected at baseline.

For unidentified households, we used a combination of three approaches interchangeably to pinpoint the location of the household: (1) a snowballing tracking approach that involved finding households by asking other members of their community where they were; (2) a GIS approach that involved using GPS coordinates from baseline; and (3) an approach that relied on using telephone numbers collected at baseline to contact missing households. The data collection teams started prioritizing these techniques on March 15, 2019. Combining these approaches helped to reduce the attrition rate. We succeeded in tracking 91.61% of the households we interviewed during the baseline data collection.

We found no difference in baseline characteristics between children in the treatment and control groups who remained in the study. That is, the reduction in sample size did not cause any differential attrition, thereby preserving the study's internal validity created through randomization. We also did not find statistically significant differences in baseline values of EGRA, EGMA, ZAT, and Oral Vocabulary scores between children who remained and children who withdrew from the sample. Table 4 below shows our test for differential attrition.

Table 4. Differential Attrition

Variables	Nonattrited		Attrited		Difference Test			Std. Mean Difference
	Mean	*N1*	Mean	*N2*	Diff	SE	*p*-Value	
Treatment	0.48	1,700	0.46	165	-0.02	0.07	0.81	-0.03
Zambian Achievement Test (% correct)	0.44	1,700	0.46	165	0.01	0.02	0.53	0.05

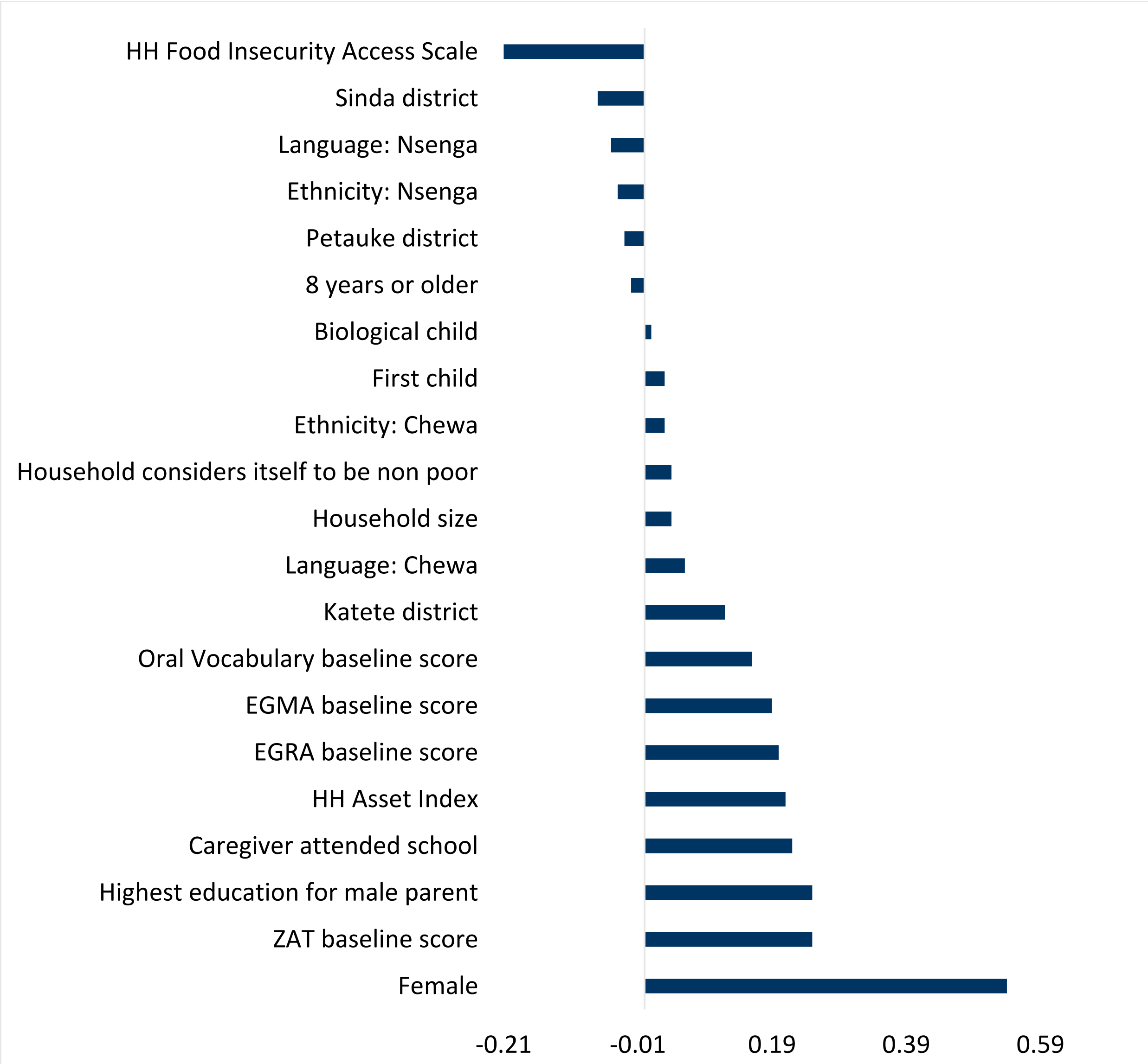
Variables	Nonattrited		Attrited		Difference Test			Std. Mean Difference
	Mean	*N*1	Mean	*N*2	Diff	SE	*p*-Value	
Early Grade Reading Assessment (% correct)	0.05	1,700	0.05	165	0.00	0.00	0.75	0.02
Early Grade Mathematics Assessment (% correct)	0.07	1,700	0.07	165	0.00	0.01	0.83	0.02
Oral Vocabulary (% correct)	0.60	1,700	0.64	165	0.03	0.03	0.29	0.11

Note. Standard errors (SE) clustered at the school level.

Impacts on School Enrollment and Program Take-Up—An analysis of school enrollment found sizeable, and statistically significant, impacts on school enrollment. In the treatment areas, 62% of the eligible children enrolled in school, while 55% of the children in control areas enrolled in school. We estimated statistical tests on the differences across children who enrolled and those who did not and found several significant differences between the two groups, which we report in Figure 4.

First, children who scored slightly higher on EGRA, EGMA, ZAT, and the Oral Vocabulary assessment at baseline were more likely to enroll in school. Second, girls were more likely than boys to enroll in school. Third, the results suggested that children in households with higher socio-economic status were more likely to enroll in school. Parents of children who enrolled in school had higher levels of education, and their households had higher levels of assets and food security compared with parents and households of children who did not enroll in school. We did not, however, find large differences in age, geography, birth order, household size, ethnicity, and language between enrolled and nonenrolled children. Tables A-1 through A-3 in Appendix A of this report provide detailed statistics regarding these samples.

Figure 4. Standardized Mean Differences Between Enrollees and Nonenrollees



We also found several statistically significant differences between children enrolled in IN schools and children who lived in the treatment areas but did not enroll in IN schools, suggesting that treatment effects on the treated applied to a sample with different characteristics than ITT effects. We estimated a multivariate regression model to predict the likelihood of enrolling in IN school for children residing in treatment areas based on demographic and household characteristics as well as baseline performances on tests. We report these results in Table 5. Overall, the results indicate that children enrolled in IN schools were significantly more likely to be female, had higher achievement levels on tests at baseline, had parents/caregivers with higher education levels, and resided in households with higher asset ownership. A multivariate logit regression model showed that the odds of enrollment in

IN schools were almost 209% higher for female children compared with male children ($p < 0.01$) and 71% higher for children from households where the caregiver had attended school ($p < 0.01$). Odds of enrollment also were significantly positively correlated with baseline scores on EGRA and EGMA ($p < 0.05$).

Table 5. Enrollment in IN Schools in IN Catchment Areas

Enrollment in IN Schools	(1) Logit Coefficients	(2) Odds Ratios
Index child was female	1.128***	3.088***
	(0.201)	(0.620)
Index child was age 6 at baseline	-0.264	0.768
	(0.209)	(0.160)
Index child was age 7 at baseline	0.136	1.146
	(0.217)	(0.249)
Index child was age 8 at baseline	-0.143	0.866
	(0.209)	(0.181)
Index child was a biological child	0.065	1.067
	(0.185)	(0.197)
Household size	0.068	1.070
	(0.046)	(0.049)
Highest level of education achieved by adult male member	-0.010	0.990
	(0.023)	(0.023)
Caregiver had attended school	0.536***	1.710***
	(0.208)	(0.356)
Petauke District	-0.109	0.897
	(0.525)	(0.470)
Sinda District	-0.491	0.612
	(0.450)	(0.275)
Zambian Achievement Test (percentage correct)	0.149	1.160
	(0.458)	(0.532)
Early Grade Reading Assessment (percentage correct)	4.404*	81.810*
	(2.330)	(190.578)

Enrollment in IN Schools	(1) Logit Coefficients	(2) Odds Ratios
Early Grade Mathematics Assessment (percentage correct)	1.591*	4.910*
	(0.914)	(4.490)
Oral Vocabulary (percentage correct)	0.137	1.147
	(0.313)	(0.359)
Index child ethnicity: Other	0.057	1.058
	(0.621)	(0.658)
Index child ethnicity: Tumbuka	-0.624	0.536
	(1.141)	(0.612)
Index child language: Nsenga	-0.347	0.707
	(0.299)	(0.212)
Index child is the first child	-0.001	0.999
	(0.589)	(0.589)
Index child is the second child	0.130	1.139
	(0.513)	(0.584)
Index child is the third child	-0.103	0.902
	(0.587)	(0.530)
Index child is the fourth child	0.077	1.080
	(0.587)	(0.634)
HH Food Insecurity Access Scale	-0.005	0.995
	(0.016)	(0.016)
HH Asset Index	0.319***	1.376***
	(0.094)	(0.130)
Baseline HH distance from school (Km)	-0.140	0.869
	(0.107)	(0.093)
Observations	794	794
Pseudo R^2	0.097	0.097

Note. Standard errors clustered at school level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Impact Evaluation Results

Primary Outcomes—ZAT, Reading, Mathematics, and Oral Vocabulary

ITT Effects on Complete Sample

Compared with baseline, we found some increases in the difference in test scores of treatment and control children 14 months after the program. Figure 5 shows the mean percentage scores (as % correct) across treatment and control groups in the two rounds of testing along with 95% confidence intervals.

Results from the formal ANCOVA framework showed statistically significant and positive effects of residing in the IN catchment areas on all primary test scores (see Table 6). The findings indicate that residing in IN catchment areas (treatment assignment) led to a 0.16 standard deviation or 3.1 percentage points increase in ZAT scores ($p < 0.05$), a 0.40 standard deviation or 3.5 percentage points increase in EGRA scores ($p < 0.01$), a 0.22 standard deviation or 4.9 percentage points increase in EGMA scores ($p < 0.05$), and a 0.25 standard deviation or 6.0 percentage points increase in Oral Vocabulary scores ($p < 0.01$). [7, 8]

These results appear to suggest a larger effect on EGRA than on EGMA outcomes, even though we found similar impact estimates when examining effects in actual scale (percentage correct). We found impact estimates of 3.5 percentage points for EGRA and 4.9 percentage points for EGMA outcomes, indicating greater variation in reading compared with mathematics achievement among the population. Specifically, the EGRA scores for children in the IN catchment areas were slightly more skewed to the right, while the distribution of EGMA scores was similar across treatment and control households.

Table 6. ITT Effects on Primary Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ZAT- % Score	ZAT- SMD	EGRA- % Score	EGRA-SMD	EGMA- % Score	EGMA-SMD	OV- % Score	OV-SMD
Treatment	0.031***	0.158***	0.035***	0.404***	0.049***	0.219***	0.060***	0.251***
	(0.011)	(0.056)	(0.007)	(0.083)	(0.015)	(0.065)	(0.013)	(0.053)
Unadjusted *p*-value	0.006	0.006	0.000	0.000	0.001	0.001	0.000	0.000

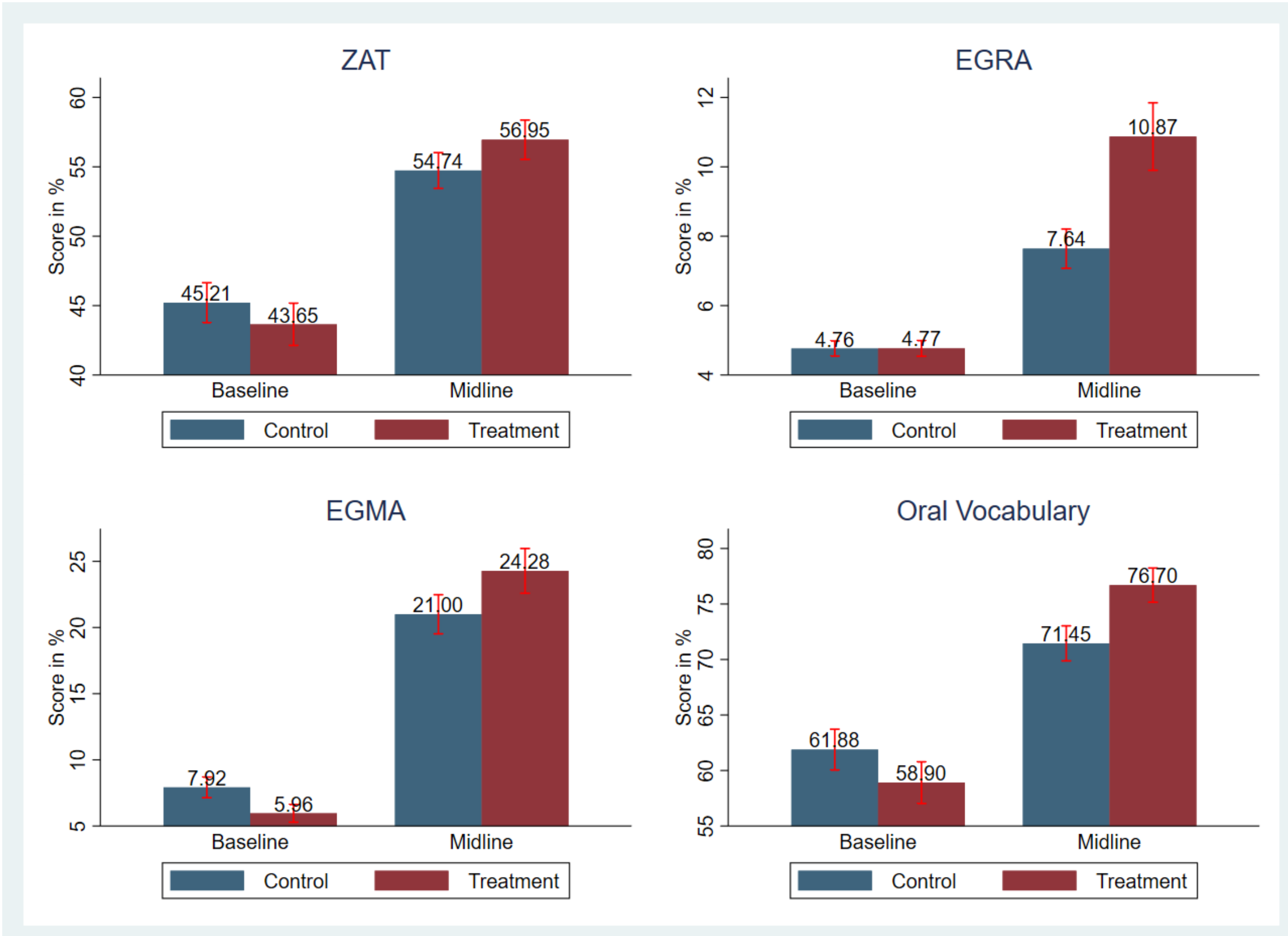
[7] In addition to the overall scores, we also found that residing in Impact Network catchment areas increased the likelihood of obtaining a positive score (or a non-zero score) on EGRA by 4.7 percentage-points, on EGMA by 1.7 percentage-points, and on Oral Vocabulary by 2.2 percentage points. See Table D-1 in Appendix D of this report.

[8] All results also show statistically significant effects after adjusting standard errors based on randomization inference as shown in the adjusted *p*-values.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	**ZAT- % Score**	**ZAT- SMD**	**EGRA- % Score**	**EGRA-SMD**	**EGMA- % Score**	**EGMA-SMD**	**OV- % Score**	**OV-SMD**
Adjusted *p*-value using randomization inference	0.007	0.007	0.000	0.000	0.001	0.001	0.000	0.000
Observations	1,688	1,688	1,688	1,688	1,688	1,688	1,688	1,688
R-squared	0.054	0.054	0.056	0.056	0.068	0.068	0.047	0.047
Control group mean	0.548		0.0766		0.210		0.714	

Note. Standard errors clustered at school level and reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.
OV = Oral Vocabulary

Figure 5. Mean Primary Test Scores (% Correct Responses) Across Treatment and Control Groups

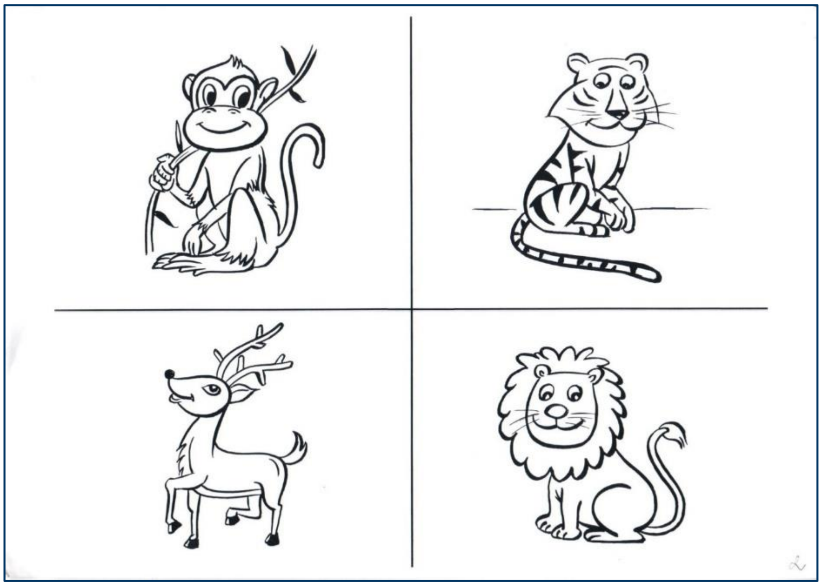


Note. N = 1,700. Across all populations, we found that the difference in scores of treatment and control children became larger and more significant at midline compared with baseline.

Below, we provide several examples from the four tests to show how ITT effects translated into differences in learning outcomes.

1. ZAT: On average, children living in the IN catchment areas scored 3.1 percentage points higher than other children on tasks such as:
    a. Recognizing the letters FI among a list of letter groups that included IP, TT, TI, and FI.
    b. Recognizing and pointing to the letter R when shown a group of pictures that included a lion, a giraffe, a rhinoceros, and the letter R.
    c. Pointing to a picture of an object that began with the same sound as another object among a group of objects.
2. EGRA: On average, children living in IN catchment areas scored 3.5 percentage points higher than other children on tasks such as:
    a. Answering questions based on reading a story about a friendship between a child and a snake.
    b. Determining where a sentence began, in which direction they were to read, and where to go when they reached the end of a line.
    c. Answering questions based on listening to a story about a child going to school for the first time during the rainy season.
3. EGMA: On average, children living in IN catchment areas scored 4.9 percentage points higher than other children on tasks such as:
    a. Recognizing numbers such as 9, 97, or 468.
    b. Identifying which number was missing from a set of patterns (for example, from a pattern of 10, 20, …, and 40, the child was asked to identify the missing number).
    c. Adding two numbers, such as 67 + 25.
4. Oral Vocabulary: Children living in IN catchment areas scored 6.0 percentage points higher on tasks like listening to a word given to them and pointing to a picture in a booklet that best represented that word. Figure 6 is an example given to children participating in this study: An enumerator may, for instance, have said the word "lion" and then asked a child to point to the image of a lion.

Figure 6. Example of a Question From the Oral Vocabulary Test



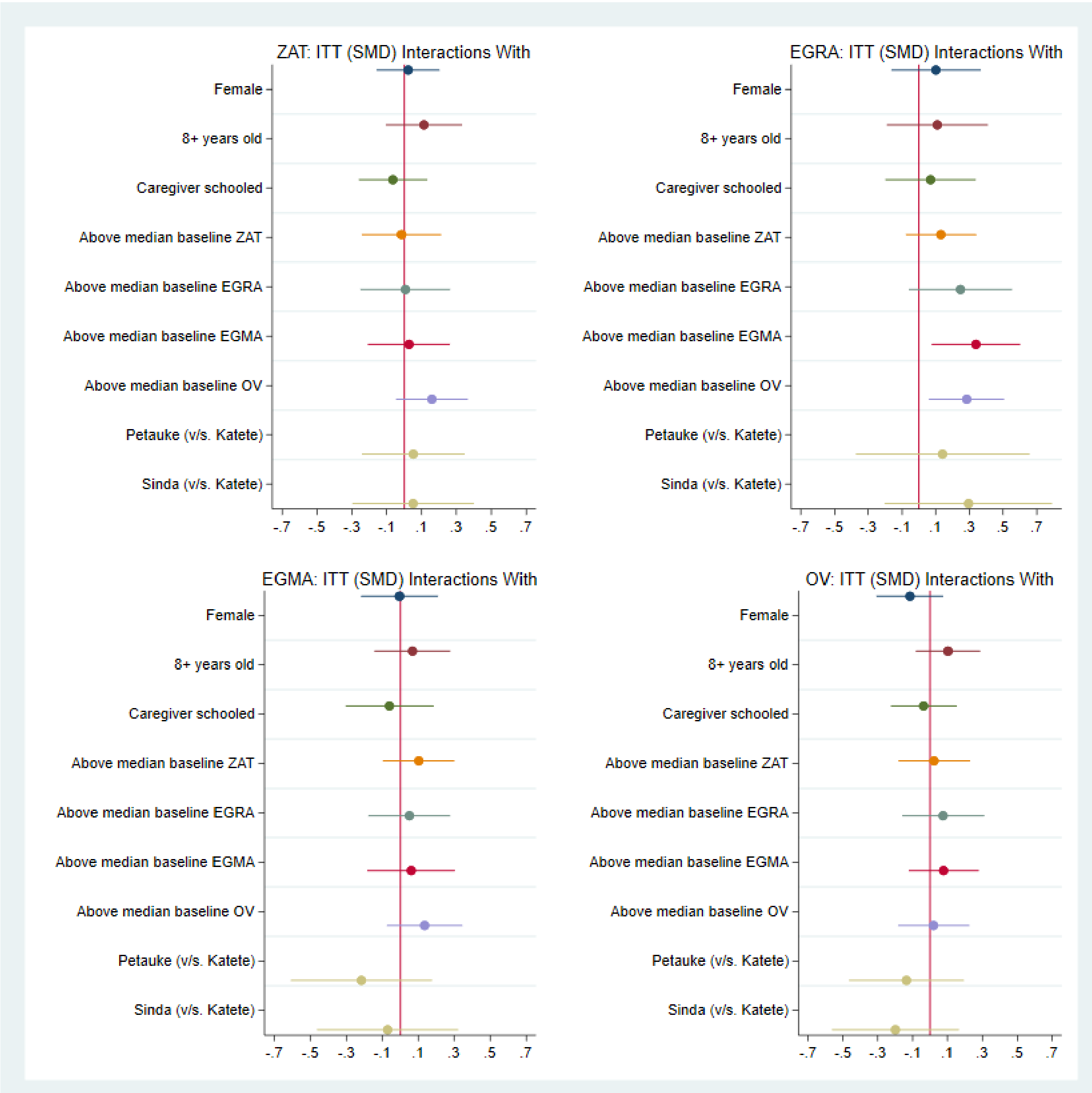
Heterogeneous Effects

We did not find substantial variation in the effects of residing in IN catchment areas by subgroups. Figure 7 shows ITT effects from different regressions including interactions between treatment assignment and: (a) child gender (female versus male), (b) child age at baseline (age 6 years or 7 years versus age 8 years or 9 years), (c) caregiver's education (whether the caregiver ever attended school), (d) household district (Katete, Sinda, or Petauke), (e) child score on baseline ZAT (above versus below median score), (f) child score on baseline EGRA (above versus below median score), and (g) child score on baseline EGMA (above versus below median score). The figure shows that almost all interaction terms were statistically insignificant, suggesting that the eSchool 360 program produced similar benefits for each of the subgroups in our sample. The results did suggest that residing in an IN catchment area had a larger impact on EGRA scores at midline for children who scored above median score on EGMA and Oral Vocabulary at baseline, but these results were only marginally significant ($p<0.10$) after correcting for multiple comparisons, as shown in Tables B-1 to B-8 in Appendix B of this report. In general, we need to exercise some caution in interpreting these results because we may not have adequate statistical power to detect small heterogeneities in the effect sizes with sufficient precision.

TOT

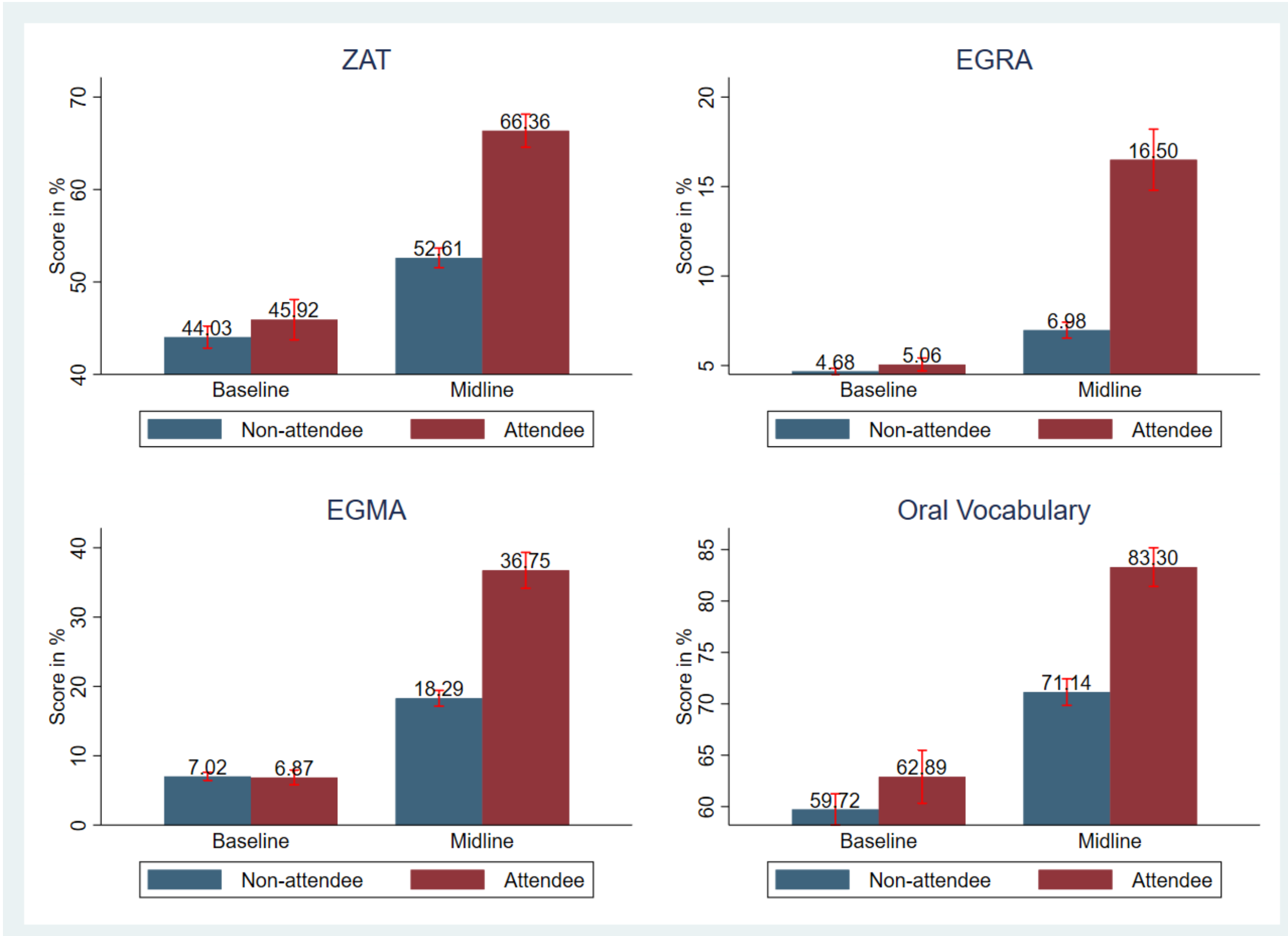
To estimate the TOT effects, we first estimated the impact of attending school at least three times in the week before the survey as an indicator of consistent attendance at school. Figure 8 shows the difference in test scores from baseline to midline for these children, referred to as IN attendees, compared with non-attendees. We find substantially larger increases in test scores for attendees across all four tests compared with non-attendees.

Figure 7. Heterogeneous ITT Effects



Note. OV = Oral Vocabulary

Figure 8. Mean Primary Test Scores (% Correct Responses) Across IN Attendees and Non-Attendees



Note. N = 1,700. The term "Attendees" refers to children who were reported to have attended an IN school at least three times in the last week. The term "Non-attendees" refers to all others.

We also present results from estimating the impact of having ever enrolled in an IN school over the year prior to the survey. We show all TOT effects formalized in an instrumental variable framework in Table 7, where column (1) shows the relationship between IN school assignment and IN school enrollment; and columns (2) to (9) show second-stage estimates reflecting the treatment effects on the treated.

Table 7. TOT Effects on Primary Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	1st Stage	ZAT- % Score	ZAT- SMD	EGRA- % Score	EGRA- SMD	EGMA- % Score	EGMA- SMD	OV- % Score	OV- SMD
Panel A									
Treatment	0.486***								
	(0.027)								
Attended IN school at least three times last week		0.063***	0.325***	0.072***	0.834***	0.101***	0.450***	0.124***	0.516***
		(0.021)	(0.108)	(0.014)	(0.159)	(0.028)	(0.123)	(0.025)	(0.103)
Observations	1,688	1,688	1,688	1,688	1,688	1,688	1,688	1,688	1,688
R-squared	0.331	0.107	0.107	0.140	0.140	0.147	0.147	0.082	0.082
Control group mean		0.526		0.0699		0.183		0.710	
Panel B									
Treatment	0.597***								
	(0.030)								
Ever attended an IN school		0.052***	0.264***	0.058***	0.677***	0.082***	0.366***	0.101***	0.420***
		(0.017)	(0.088)	(0.011)	(0.131)	(0.022)	(0.100)	(0.021)	(0.087)
Observations	1,688	1,688	1,688	1,688	1,688	1,688	1,688	1,688	1,688
R-squared	0.435	0.095	0.095	0.122	0.122	0.130	0.130	0.079	0.079
Control group mean		0.526		0.0689		0.181		0.707	

Note. Standard errors clustered at school level and reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TOT effects showed that attending an IN school at least three times in the week before the survey had a significant positive impact on all primary test scores. Specifically, we found that that proxy for active participation in the program led to an increase of 0.32 standard deviation or 6.3 percentage points in the ZAT score, a 0.83 standard deviation or 7.2 percentage point increase in the primary EGRA score, a 0.45 standard deviation or 10.1 percentage point increase in the primary EGMA score, and a 0.52 standard deviation or 12.4 percentage point increase in the Oral Vocabulary score. We found that children residing in an IN school catchment area were 49 percentage points more likely than other children to have attended an IN school at least three times in the week.

The TOT effects remain substantial but are smaller when we use ever enrolling in an IN school as a proxy for participation in the program. We found that enrollment in IN schools led to an increase of 0.26 standard deviations or 5.2 percentage points in the ZAT score, a 0.68 standard deviation or 5.8 percentage points increase in the primary EGRA score, a 0.37 standard deviation or 8.2 percentage points increase in the primary EGMA score, and a 0.42 standard deviation or 10.1 percentage points increase in the Oral Vocabulary score. The effects are likely smaller than in the previous analysis because children who were ever enrolled in IN schools may not have consistently attended school, limiting the benefits of their enrollment. Indeed, the development economics literature suggests that enrolling in school is not always a good proxy for attending school (Alderman, Gilligan, and Lehrer, 2012).

Overall, the TOT estimates are larger than the ITT estimates because the ITT effects compared all children in IN catchment areas with other children irrespective of whether they attended an IN School. The ITT estimates are, therefore, a weighted average of effects on children who were and who were not enrolled in school, while the TOT results estimated the impact of attending an IN school (proxied by either attending school at least three times in the week before the survey or being ever enrolled in IN schools).

Equivalent Years of Schooling

We found effect sizes for EGRA and EGMA assessments that were equivalent to 1.88 additional years of education for reading and 1.03 additional years of education for mathematics if we transformed the effect sizes of 0.40 standard deviations on EGRA scores and 0.22 standard deviations on EGMA scores and extrapolated results of learning gains from Grade 1 to Grade 12 from a sample of representative low- and middle-income countries (Evans & Yuan, 2019). The equivalent years of schooling were based on the finding that 4.7 years of schooling was equivalent to a one standard deviation improvement in reading proficiency in a pooled sample of five representative low-and middle-income countries. When compared directly against the one-year learning gains in the control group in the original test scale and units, the ITT estimates are equivalent to 1.2 additional years of education for reading and 0.37 additional years (or 4.5 additional months) of education for mathematics.

Reading and Mathematics Subscores

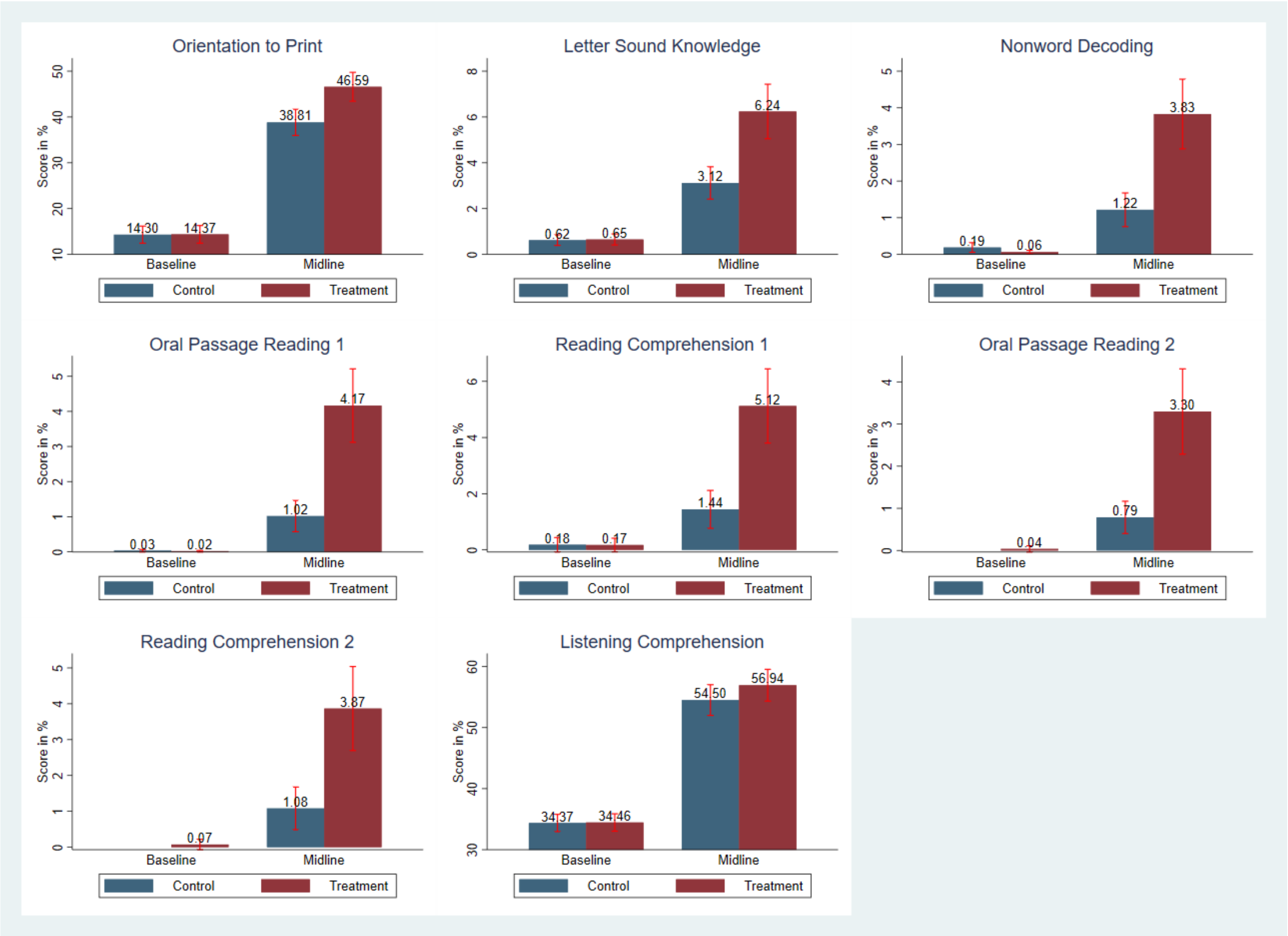
Subscores on Early Reading Assessment (EGRA)

Previous experience in Zambia suggests that a majority of first-grade students score low on their EGRA and EGMA tests, with many zero scores, indicating that we may find larger effects on the easier EGRA and EGMA subcomponents that are administered in the early stages of the assessment than on the more complex EGRA subcomponents that are administered in the later stages of the assessment. However, we found that living in IN catchment areas had a significant positive impact on all but one EGRA subcomponent, which suggests that the eSchool 360 model leads to improvements in easier and more complex EGRA subtasks. Figure 9 shows descriptive differences in mean scores from baseline for the treatment and control groups.

Columns (1) through (3) in Tables 8 and 9 show statistically significant ITT effects on seven out of eight subtasks of the EGRA assessment: orientation to print, letter sound knowledge, nonword decoding, oral passage reading (basic and higher levels), and reading comprehension (basic and higher levels). The largest effect in percentage points was for orientation to print (8.7 percentage points), while most of the other effect sizes ranged between 2 and 4 percentage points. We did not find a statistically significant effect on listening comprehension. The results were robust to correcting for multiple comparisons using the Benjamini-Hoschberg method (as shown in the adjusted *q*-value in column (3)).

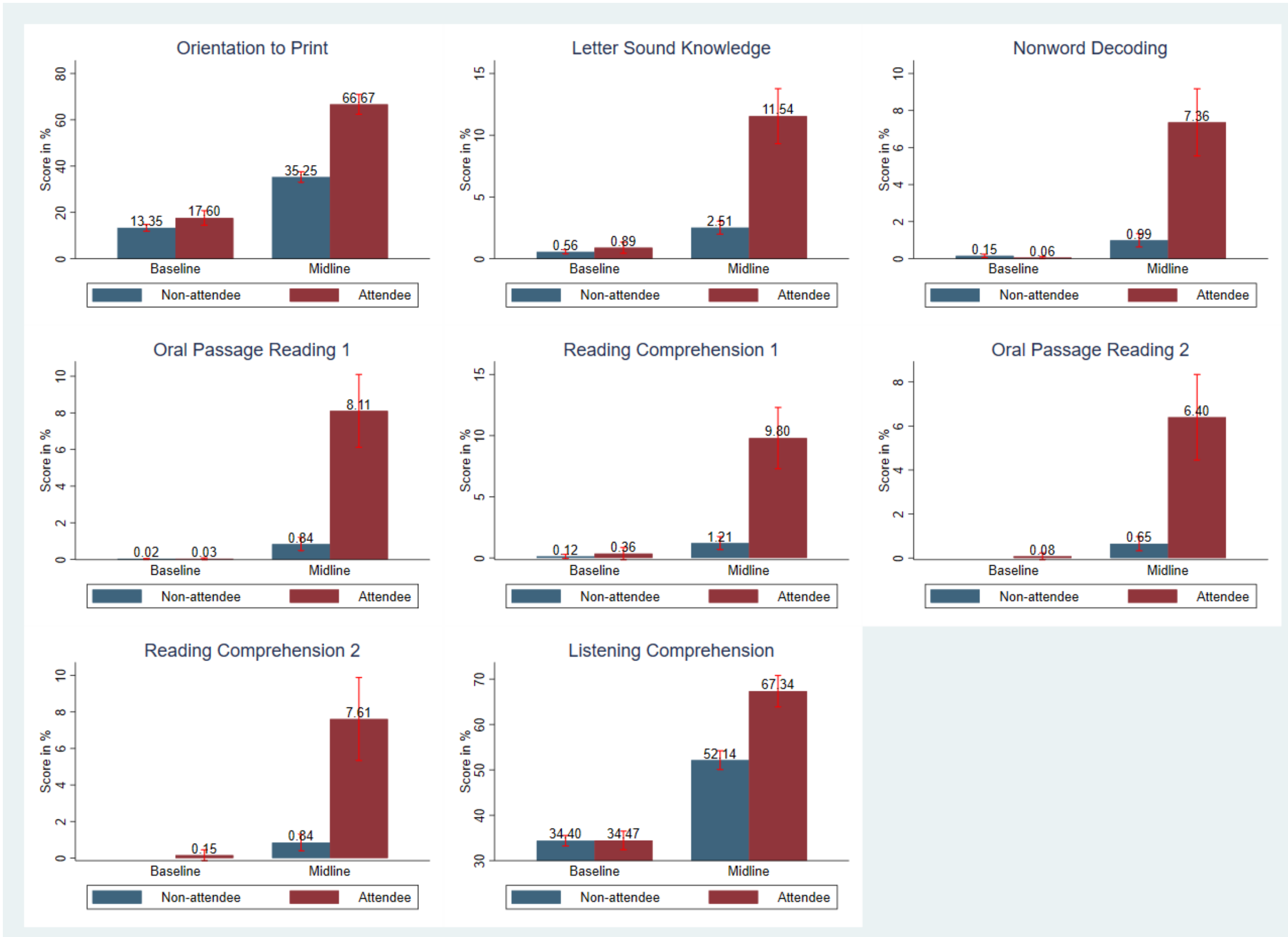
As expected, we find much larger differences when comparing children who attended IN schools at least three times in the last week compared with other children (Figure 10). Estimates of TOT effects (based on the proxy of attending school at least three times in the last week) show a strong positive impact of attending IN schools on all EGRA subtasks except listening comprehension (Tables 8 and 9, columns (4) though (6)). We found particularly large effect sizes for some subtasks—specifically, having enrolled in an IN school increases basic oral passage reading performance (level 1) by 1.1 standard deviations or 7.3 percentage points, and higher-level oral reading performance by 1 standard deviation or almost 5.8 percentage points (level 2).

Figure 9. Mean EGRA Subscores (% Correct Responses) Across Treatment and Control Groups



Note. N = 1,700.

Figure 10. Mean EGRA Subscores (% Correct Responses) Across IN Attendees and Non-Attendees



Note. N = 1,700. Attendees are children who were reported to have attended an IN school at least three times in the last week. Non-attendees include all others.

Table 8. ITT and TOT Results on EGRA Subscores (Standard Deviation)

	(1)	(2)	(3)	(4)	(5)	(6)
	ANCOVA Estimates: ITT Effects			IV 2SLS Estimates: TOT Effects		
Subsample	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value
Orientation to Print	0.202***	0.003	0.003	0.417***	0.001	0.001
	(0.064)			(0.123)		
Letter Sound Knowledge	0.305***	0.001	0.001	0.629***	0.000	0.000
	(0.086)			(0.166)		
Nonword Decoding	0.428***	0.000	0.000	0.882***	0.000	0.000
	(0.100)			(0.196)		
Oral Passage Reading level 1	0.524***	0.000	0.000	1.078***	0.000	0.000
	(0.113)			(0.216)		
Reading Comprehension level 1	0.404***	0.000	0.000	0.833***	0.000	0.000
	(0.093)			(0.191)		
Oral Passage Reading level 2	0.485***	0.000	0.000	0.999***	0.000	0.000
	(0.119)			(0.240)		
Reading Comprehension level 2	0.337***	0.000	0.001	0.695***	0.000	0.000
	(0.091)			(0.186)		
Listening Comprehension	0.068	0.229	0.229	0.140	0.217	0.217
	(0.056)			(0.114)		

Note. Standard errors clustered at school level and reported in parentheses. TOT effects are based on having attended an IN school at least three times in the last week. * $q < 0.1$; ** $q < 0.05$; *** $q < 0.01$

Table 9. ITT and TOT Results on EGRA Subscores (Percent Correct)

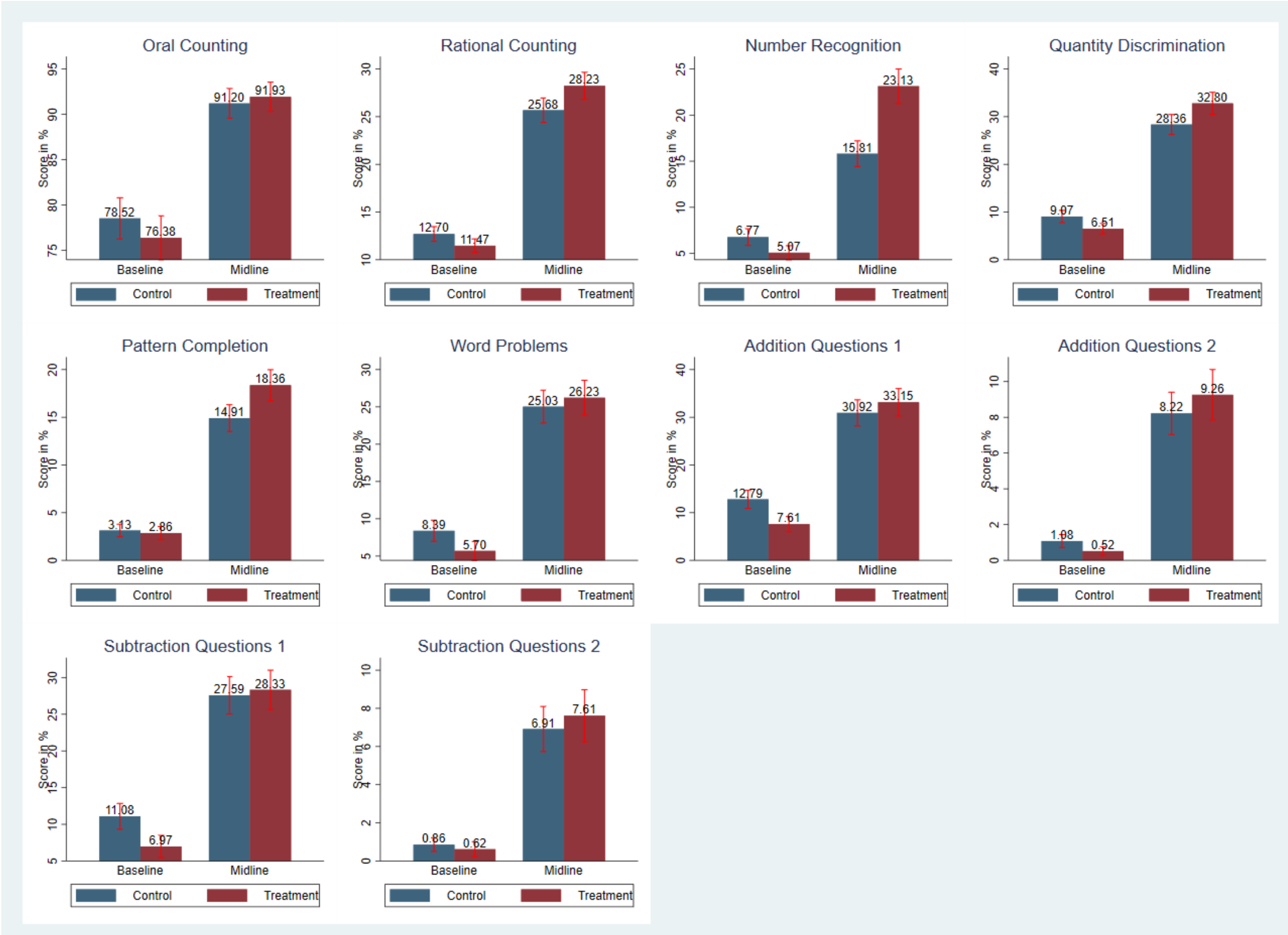
	(1)	(2)	(3)	(4)	(5)	(6)
	ANCOVA Estimates: ITT Effects			IV 2SLS Estimates: TOT Effects		
Subsample	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value
Orientation to Print	0.087***	0.003	0.003	0.181***	0.001	0.001
	(0.028)			(0.053)		
Letter Sound Knowledge	0.033***	0.001	0.001	0.067***	0.000	0.000
	(0.009)			(0.018)		
Nonword Decoding	0.030***	0.000	0.000	0.062***	0.000	0.000
	(0.007)			(0.014)		
Oral Passage Reading Level 1	0.035***	0.000	0.000	0.073***	0.000	0.000
	(0.008)			(0.015)		
Reading Comprehension Level 1	0.041***	0.000	0.000	0.085***	0.000	0.000
	(0.010)			(0.020)		
Oral Passage Reading Level 2	0.028***	0.000	0.000	0.058***	0.000	0.000
	(0.007)			(0.014)		
Reading Comprehension Level 2	0.031***	0.000	0.001	0.063***	0.000	0.000
	(0.008)			(0.017)		
Listening Comprehension	0.026	0.229	0.229	0.054	0.217	0.217
	(0.021)			(0.043)		

Note. Standard errors clustered at school level and reported in parentheses. TOT effects are based on having attended an IN school at least three times in the last week. * $q < 0.1$; ** $q < 0.05$; *** $q < 0.01$

Subscores on Early Grade Mathematics Assessment (EGMA)

Figure 11 shows descriptive differences in mean scores from baseline for the treatment and control groups. We see particularly large increases in scores on rational counting, number recognition, and pattern completion.

Figure 11. Mean EGMA Subscores (% Correct Responses) Across Treatment and Control Groups



Note. N = 1,700.

The ITT estimates showed statistically significant effects on most of the EGMA subtasks, with larger effects on easier EGMA subtasks. The effect sizes in standard deviations for EGMA were smaller than on the EGRA subtasks, even though effects in actual scale were slightly higher on EGMA subtasks, again reflecting the more positively skewed distribution of EGRA scores for children in treatment areas. Columns (1) through (3) of Table 10 show that being assigned to IN had a small but significant positive impact on rational counting, number recognition, quantity discrimination, and pattern completion. We found no statistically significant impacts on oral counting, word problems, and two levels each of addition and subtraction questions.

Again, we see much higher scores on all EGMA subtasks for the sample that reported having attended an IN school at least three times in the previous week (Figure 12). However, estimates of TOT based on the proxy of attending school at least three times in the week before the survey did not show statistically significant effects for most EGMA subtasks except rational counting (a 0.43 standard deviation increase), number recognition (a 0.84 standard deviation increase), and quantity discrimination (a 0.42 standard deviation increase). Columns 4–6 of Table 10 depict these results.

Table 10. ITT and TOT Results on EGMA Subtask Scores (Standard Deviation)

	(1)	(2)	(3)	(4)	(5)	(6)
	ANCOVA Estimates: ITT Effects			IV 2SLS Estimates: TOT Effects		
Subsample	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value
Oral Counting	0.075	0.136	0.170	0.154	0.124	0.155
	(0.050)			(0.100)		
Rational Counting	0.210***	0.000	0.001	0.433***	0.000	0.000
	(0.054)			(0.099)		
Number Recognition	0.408***	0.000	0.000	0.840***	0.000	0.000
	(0.078)			(0.144)		
Quantity Discrimination	0.205***	0.001	0.003	0.423***	0.000	0.001
	(0.060)			(0.112)		
Pattern Completion	0.222***	0.004	0.010	0.456***	0.001	0.003
	(0.074)			(0.143)		
Word Problems	0.090	0.112	0.160	0.185	0.095	0.136
	(0.056)			(0.111)		
Addition Questions (Level 1)	0.110	0.062	0.124	0.226	0.046	0.092
	(0.058)			(0.113)		
Addition Questions (Level 2)	0.119	0.085	0.141	0.245	0.068	0.113
	(0.068)			(0.134)		
Subtraction Questions (Level 1)	0.077	0.168	0.187	0.159	0.146	0.162
	(0.055)			(0.109)		
Subtraction Questions (Level 2)	0.083	0.244	0.244	0.171	0.227	0.227
	(0.070)			(0.141)		

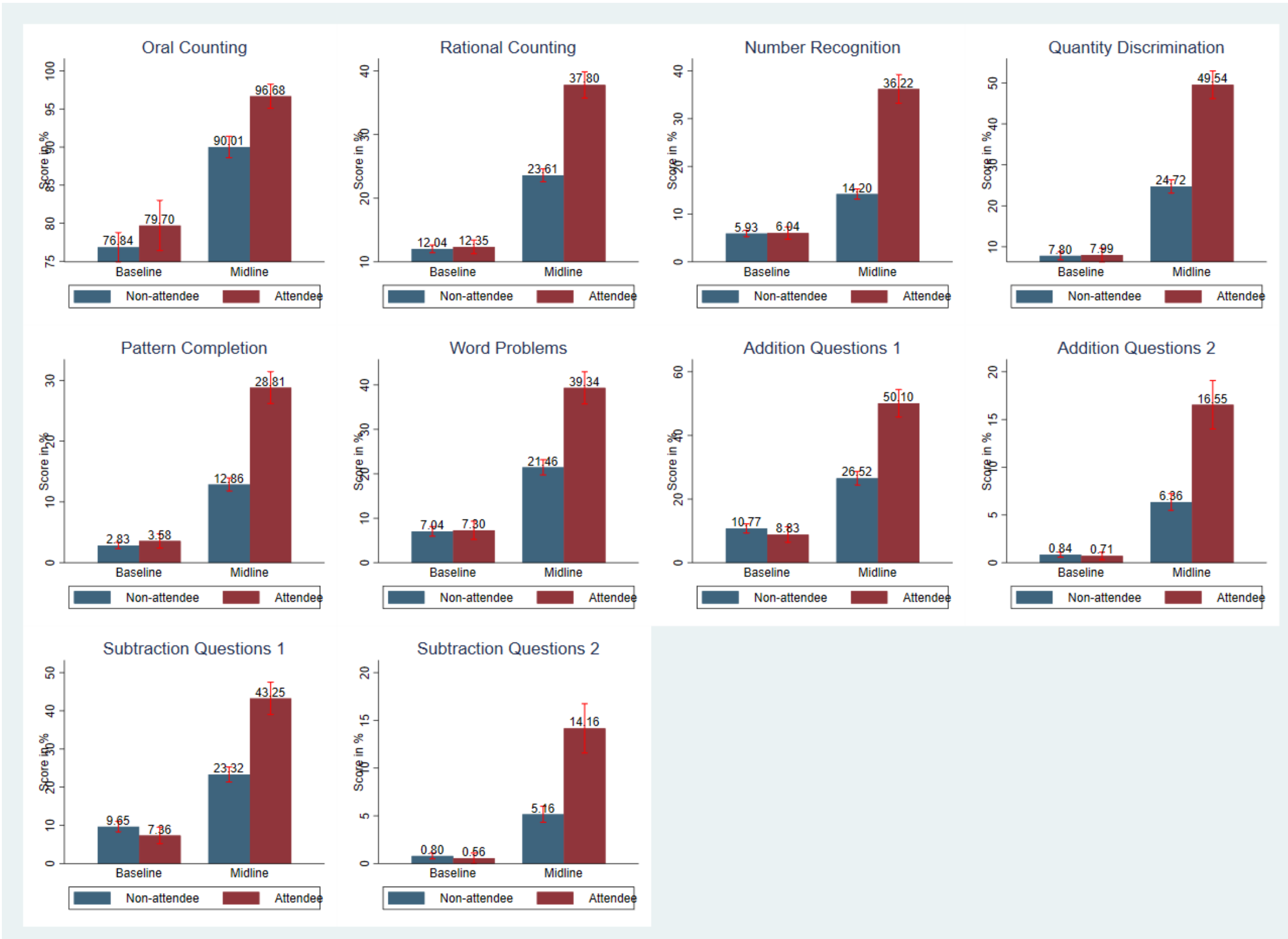
Note. Standard errors (SE) clustered at school level and reported in parentheses. TOT effects are based on having attended an IN school at least three times in the last week. * $q < 0.1$; ** $q < 0.05$; *** $q < 0.01$

Table 11. ITT and TOT Results on EGMA Subtask Scores (Percent Correct)

	(1)	(2)	(3)	(4)	(5)	(6)
	ANCOVA Estimates: ITT Effects			IV 2SLS Estimates: TOT Effects		
Subsample	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value	Coef. (SE)	Unadjusted *p*-Value	Adjusted *q*-Value
Oral Counting	0.019	0.136	0.170	0.038	0.124	0.155
	(0.012)			(0.025)		
Rational Counting	0.041***	0.000	0.001	0.084***	0.000	0.000
	(0.010)			(0.019)		
Number Recognition	0.087***	0.000	0.000	0.178***	0.000	0.000
	(0.017)			(0.030)		
Quantity Discrimination	0.065***	0.001	0.003	0.135***	0.000	0.001
	(0.019)			(0.036)		
Pattern Completion	0.047***	0.004	0.010	0.098***	0.001	0.003
	(0.016)			(0.031)		
Word Problems	0.030	0.112	0.160	0.062	0.095	0.136
	(0.019)			(0.037)		
Addition Questions (Level 1)	0.046	0.062	0.124	0.094	0.046	0.092
	(0.024)			(0.047)		
Addition Questions (Level 2)	0.021	0.085	0.141	0.044	0.068	0.113
	(0.012)			(0.024)		
Subtraction Questions (Level 1)	0.030	0.168	0.187	0.061	0.146	0.162
	(0.021)			(0.042)		
Subtraction Questions (Level 2)	0.015	0.244	0.244	0.031	0.227	0.227
	(0.013)			(0.025)		

Note. Standard errors (SE) clustered at school level and reported in parentheses. TOT effects are based on having attended an IN school at least three times in the last week. * $q < 0.1$; ** $q < 0.05$; *** $q < 0.01$

Figure 12. Mean EGMA Subscores (% Correct Responses) Across IN Attendees and Non-Attendees



Note. N = 1,700. Attendees are children who were reported to have attended an IN school at least three times in the last week. Non-attendees include all others.

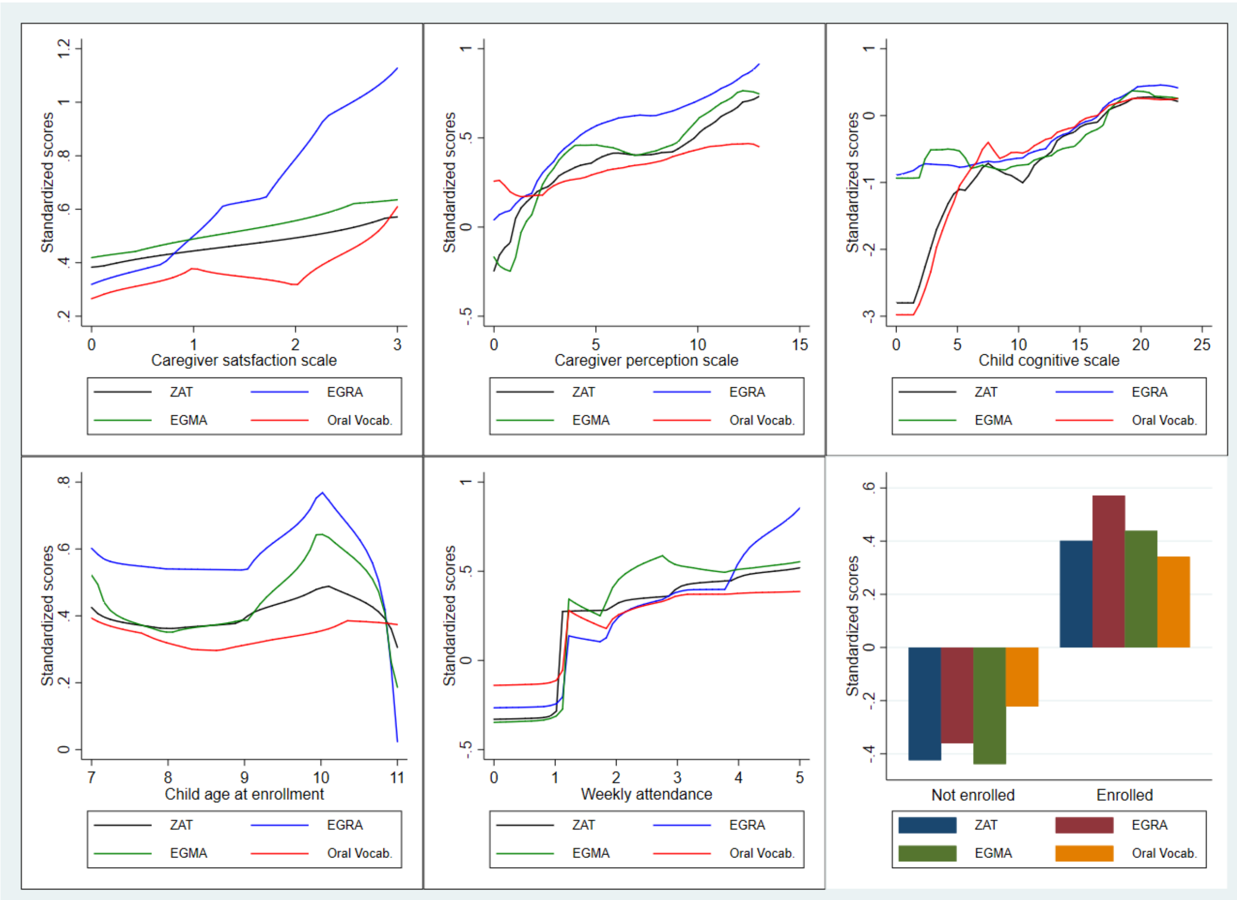
In line with our quantitative finding that the IN program had statistically significant positive impacts on literacy, qualitative data showed perceived improvements in children's literacy according to parents, teachers, and IN staff. Parents and program staff noted that children at IN schools read better than their counterparts attending government-run schools and started to read earlier (in Grade 1). One parent from Kabanga School commented to this end: *"…when you take the children in a government school…they have difficulties in knowing how to read; but with this* [IN] *school they easily know how to read."* Children, for their part, shared how they learned to read through activities involving the projector. When asked why they liked the projector, one student from Bvunda School said, "*We learn how to read through it. We get words."* Classroom observations confirmed that children did various literacy-related activities using technology, including listening to stories via the projector and reading stories in their tablets. Interestingly, however, respondents did not report perceived changes in numeracy resulting from the IN program.

Mechanisms of Impact

This section examines impacts on intermediate outcomes to assess potential mechanisms that drove the impacts on EGRA, EGMA, ZAT, and Oral Vocabulary assessments. For this analysis, we examined impacts on intermediate outcomes followed by an analysis of the correlation between intermediate outcomes and EGRA, EGMA, ZAT, and Oral Vocabulary scores.

We first assessed the program impacts on school enrollment and attendance to determine whether increased schooling may have contributed to the positive impacts. Second, we assessed the effects on parents' perceptions of school and education quality by estimating impacts on 4-point Likert-scale questions related to parents' general perceptions on the quality of education as well as to more specific attitudes related to IN's activity-based curriculum, the use of technology in the classroom, and teachers' pedagogical practices. We examined program impacts on children's level of engagement in the classroom, the time devoted to activity-based learning activities, the use of technology in the classroom, and children's interaction with teachers. Third, we estimated impacts on child development, which we measured by using an index constructed based on answers of parents about the behavior of their children. Figure 13 plots correlations between each of these factors and test scores across the four assessments; and Tables C-1 and C-2 of Appendix C show results from multivariate regressions depicting the correlations between multiple factors and test scores.

Figure 13. Intermediate Outcomes and Primary Test Scores



School Enrollment and Attendance

We found that living in IN catchment areas increased the likelihood of school enrolment by almost 8 percentage-points. Figure 13 shows that school enrollment is positively correlated with test scores, suggesting that increased likelihood of enrollment was likely an important mechanism for improvements in learning outcomes due to the eSchool 360 model. We also found that residing in IN area led to an increase of 0.36 days of weekly attendance (or 1.6 days in a month), on average (Table 12). Figure 13 shows that learning outcomes on all four tests jump beyond one day of weekly attendance but increase at a modest pace beyond that. This may be indicative of a significant—although small—mediating effect of school attendance.

Table 12. ITT Effects on School Enrollment and Attendance

VARIABLES	(1) Enrolled (yes/no)	(2) Enrolled (SMD)	(3) Number of Days Attended	(4) Number of Days Attended (SMD)	(5) Age at Enrollment	(6) Age at Enrollment (SMD)
Treatment	0.079**	0.159**	0.358**	0.162**	-0.096**	-0.094**
	(0.038)	(0.075)	(0.158)	(0.071)	(0.038)	(0.037)
Observations	1,688	1,688	1,688	1,688	979	979
R-squared	0.021	0.021	0.024	0.024	0.800	0.800
Control group mean	0.545		1.915		9.044	

Note. Standard errors clustered at school level and reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Results from interviews and focus groups also showed that parents, teachers, and program staff noted high student attendance and enrollment in IN classes. A member of IN's staff noted that the "average student attendance was about 90%." A few respondents mentioned that children were motivated to attend class due to interest in using tablets and technology in the classroom. A teacher explained, *"I think we have the iSchool, where they* [students] *come to school and learn using the projector and tablets. I think they feel it's a plus because they feel good to come to school and learn the new method of teaching* [so] *they attend classes."* Program staff felt that attendance and enrollment was higher in IN classes compared with grades that did not offer the eSchool 360 program.

Parents and program staff mentioned that some parents were less motivated to enroll their children in classes if they did not receive the eSchool 360 intervention. According to respondents, IN transitioned from implementing the program in Grade 1 to Grade 2 during the 2018–19 school year. As a result, an IN staff explained that "some parents were hesitant to enroll children in a nonprogram class." This was especially relevant to Grade 1, which previously had benefited from the eSchool 360 program. Several parents expressed concern relative to school quality and cost in the absence of the of eSchool 360 intervention.

While IN staff agreed that overall student attendance was "pretty high," they acknowledged seasonal trends in attendance and indicated attendance was lower during rainy season. IN staff reported that in the rainy season, some children were still working on their family farms, herding cattle (boys only) or caring for younger siblings while parents worked on the farm, or were unable to cross rising rivers due to rain.

Interestingly, we found that age at enrollment decreased significantly because of the introduction of the IN Schools. Table 12 shows that students in the IN catchment areas were, on average, 0.09 years younger than students in control schools. This difference is statistically significant and indicates that parents may have decided to send children to school at an earlier age because of improved perceptions of education quality or because of the zero costs of attending IN schools. However, the decreased age at enrolment likely did not contribute to improvements in learning outcomes, considering that age at enrollment did not show a monotonically positive correlation with learning outcomes (Figure 13).

Child Development

Although we found significant correlations between child development (as measured through a child cognitive skills scale) and learning outcomes (Figure 13), the ITT estimates did not show statistically significant effects of the eSchool 360 model on child development outcomes (Table 13). All point estimates were positive, but the impact estimates were not statistically significant. This finding indicates that the positive impacts on EGRA, EGMA, ZAT, and Oral Vocabulary assessments were likely not driven by improvements in child development.

Table 13. ITT Effects on Other Intermediate Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Child Developmen t Scale	Child Developmen t Scale (SMD)	Caregiver Satisfaction Scale	Caregiver Satisfaction Scale (SMD)	Caregiver Perception/ Engagemen t Scale	Caregiver Perception/ Engagemen t Scale (SMD)
Treatment	0.242	0.084	0.299**	0.284**	0.845***	0.316***
	(0.174)	(0.061)	(0.121)	(0.114)	(0.239)	(0.089)
Observations	1,688	1,688	878	878	878	878
R-squared	0.037	0.037	0.047	0.047	0.033	0.033
Control group mean	17.94		1.189		7.729	

Note. Standard errors clustered at school level and reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Parents' Satisfaction With School and Education Quality

We found statistically significant effects on parents' perceptions of school and education quality. Table 13 shows that caregivers in the IN catchment areas reported a 0.30-point higher satisfaction with education than caregivers in the control group ($p < 0.05$). We based these impact estimates on interviews with caregivers of children who enrolled in any school in the past or in the current year.

We also found positive correlations between parents' perceptions of school education quality and learning outcomes, suggesting that improvements in the quality of education may have contributed to the positive effects of the eSchool 360 model on EGRA, EGMA, ZAT, and Oral Vocabulary assessments (Figure 13).

Parents, teacher supervisors, students, and program staff shared the belief that IN teachers provided students with high-quality education. Respondents associated the provision of quality education with knowledgeable teachers who used innovative teaching methods, constant teacher attendance, teacher follow-up when students were absent, and improvements in student performance.

Qualitative results showed that parents believed that IN teachers were well-trained and able to apply effective teaching methods including the use of technology in the classroom. Several parents mentioned that IN teachers benefited from trainings that improved their teachings skills, with one parent noting, *"Impact teachers go for refresher courses frequently. So from that Impact has helped our teachers to become very educated teachers."* Parents also appreciated the use of technology in the classroom and felt that it enhanced the learning process. One parent explained how teachers used tablets to improve lesson delivery: *"What I saw in the way they use it* [the tablet]*—it has lessons in math. When they start teaching, the children follow. There are different types of lessons. And that's why the program should not end here but should continue, because the children are learning well."* Several teacher supervisors also noted that trainings increased teachers' capabilities and confidence in the classroom.

iSchool tablets.

Parents, students, teacher supervisors, and program staff agreed that IN teachers were always present in class. Respondents noted that if a teacher was absent, another teacher would cover the class. One parent explained this practice: "*She* [the teacher] *comes daily, but when she is not feeling well she calls another teacher here to come and teach the children."* Teacher supervisors and program staff agreed with this perception and noted that strong monitoring practices encouraged teachers to attend class regularly. Program staff explained that teacher supervisors and operations managers regularly monitored teacher attendance. In addition, teachers were required to bring medical notes explaining any absences. Program staff also mentioned that constant teacher attendance differentiated IN from government schools in which teachers were frequently absent.

Parents, teacher supervisors, and program staff believed that improvements in student performance demonstrated that teachers provided high-quality education to students. Respondents frequently noted that students enrolled in eSchool 360 classes had strong literacy skills compared with students enrolled in similar or higher grades in government schools. One parent explained this difference: "*The children at this school learned how to read from Grade 1; but you find a child who is in Grade 4 there* [in a government school] *but does not know how to read*." Teacher supervisors and program staff reported similar stories of students learning to read from an earlier age when enrolled in IN classes.

Parents' Perceptions of School Engagement

We found statistically significant effects of the eSchool 360 model on a parental engagement scale, but it was unclear whether these impacts contributed to increases in learning outcomes because parental engagement and parental satisfaction are highly correlated with each other. Table 14 shows that parents reported statistically significantly higher engagement levels in IN catchment areas than in control areas. While parental engagement was found to have a significant positive correlation with learning outcomes at a bivariate level (Figure 13), this relationship was not robust to controlling for parental satisfaction and other factors, as shown in Table C-2 in Appendix C, possibly because of the high correlation with parental satisfaction.

Qualitative results suggested that parents valued the fact that teachers consulted with them when a student was absent. One parent recounted frequent communication with her child's teacher whenever the child was absent from school: *"They tell us, they call us and they ask us why the child isn't coming to school."* Teachers also mentioned the protocol that they followed when students were absent, with one teacher explaining, *"If I see that the child has remained absent, I go to the parents to visit to see what is happening. Even if the child has been absent because of sickness, I'll go visit them to see how the child is feeling."* Parents mentioned that active communication with teachers demonstrated IN teachers' strong commitment to teaching.

Secondary Outcomes

This section presents program impacts on secondary outcomes, including parents' aspirations, food security, and education expenditures. First, we estimated effects on parents' aspirations with respect to their children's education and marriage. To measure parental aspirations, we asked parents for the bride prices they expected to receive for their daughters, the level of education they would like their children to attain, and their children's preferred ages at marriage. Second, we assessed program impacts on food security and education expenditures.

Aspirations About Children's Education and Marriage

We found no evidence for effects of the eSchool 360 model on parental-level aspirations. Table 14 shows that we did not find statistically significant effects on the probability parents assigned to their children studying beyond the senior secondary level (beyond 12th grade), their children's preferred age at marriage, or expected bride prices for their daughters.

Table 14. ITT Effects on Caregivers' Aspirations and Household Food Security

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Ideal Age for Marriage		Child Aspired to study beyond 12th Grade		Expected Bride Price		Household Food Security		Household School-Related Expenditure	
VARIABLES	Years	SMD	Yes/No	SMD	Zambian Kwacha	SMD	Index scale	SMD	Zambian Kwacha	SMD
Treatment	-0.081	-0.019	-0.023	-0.046	-490.963	-0.105	1.129**	0.201**	-13.878**	-0.161**
	(0.249)	(0.059)	(0.038)	(0.077)	(346.806)	(0.074)	(0.479)	(0.085)	(6.160)	(0.072)
Observations	1,625	1,625	1,619	1,619	696	696	1,109	1,109	1,109	1,109
R-squared	0.042	0.042	0.026	0.026	0.033	0.033	0.048	0.048	0.026	0.026
Control group mean	25.98		0.554		2849		11.91		55.72	

Note. Standard errors (SE) clustered at school level and reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Household Food Security and School-Related Expenditure

Interestingly, the findings demonstrated positive effects of the eSchool 360 model on the food security of households residing in IN catchment areas. We found that treatment households scored 1.13 points or 0.20 standard deviations higher on a food security scale than control households (Table 14). There are two potential reasons that we will investigate further during endline data collection: (1) The effects on food security may have been caused by increased food expenditures enabled by reductions in education expenditures due to the free education offered by IN (results shown in columns 9 and 10 of Table 14); and (2) The impact may have been channeled through an increase in school enrollment if any nutrition initiatives were implemented in schools in our catchment areas. These factors will be explored in more detail in the next wave of data collection. However, we already found some evidence for statistically significant reductions in education expenditures caused by the eSchool 360 model.

Process Evaluation Results

Fidelity of Implementation

KIIs, FGDs, and direct observations revealed that the IN program was implemented with a high level of fidelity. Students, parents, IN staff, and teachers themselves reported that IN teachers were regularly present at school, used tablets and projectors consistently, employed participatory learning approaches such as group works, and followed up with students when they were absent. Regarding teacher attendance and follow-up with absent students, one Kabanga School parent stated, *"The teacher is very good. Any time the children come to school there is no time they come back* [and say] *that the teacher is not there. When the child is absent from school the teacher makes a follow-up to us parents to get the reason the child is absent."* Qualitative respondents also confirmed that IN schools consistently had the prescribed number of tablets—seven pupil tablets and two teacher tablets.

Classroom observations showed IN teachers using active, participatory pedagogical approaches as they were trained to do, such as putting students into small groups and inviting them to actively participate in the lesson. For example, teachers called on individual students to come to the board to identify different types of punctuation during one observation. Respondents also confirmed that IN teachers closely adhered to the prescribed curriculum. Teachers followed a curriculum map that showed which lessons to teach and when, and they viewed lessons in the tablet and modified them to match the level of their learners. IN staff reported that they were fairly strict with newer teachers and asked them to adhere closely to what was in the tablet; gradually, however, teachers were given more flexibility to adapt.

One notable modification to the program was made at the beginning of 2019, when IN staff added 30 minutes of literacy-focused activities and 30 minutes of numeracy-focused activities to the daily lesson plan. These two 30-minute sessions replaced creative and technology studies lessons and were intended to prioritize learning to read, write, and count. Given that children were in school for only four hours per day, IN staff reported wanting to place greater emphasis on literacy and numeracy to accelerate learning in these two areas.

Variation by Geography, Culture, and Season

Cultural and geographic variation across the three IN districts was somewhat limited given that all program districts are in Zambia's Eastern Province, but Petauke is larger and more expansive than the other districts. It is not surprising, then, that children in Petauke reported traveling the greatest amount of time to reach school (25 minutes each way, on average) compared with Katete (17 minutes) and Sinda (11 minutes).

During interviews, IN staff shared their perception that girls attended school more frequently than boys, and our quantitative findings confirmed this. Of the IN students who attended all five of the previous school days, 57.4% were girls and 42.6% were boys. Of the IN students who did not attend school any of the five previous school days, 57.5% were boys and 42.5% were girls. IN staff also reported during interviews that students who walked more than one hour to reach school attended less regularly, and we saw quantitative evidence that the length of time required to reach school may have affected attendance. Of the children enrolled in IN schools who reported they did not attend school at all during the previous week, 83.1% lived more than 30 minutes away from their school. Conversely, of the IN children who reported attending school all five of the previous school days, 71.1% lived less than 30 minutes from the school. These data suggest that longer distances to school may have adversely affected attendance at IN schools.

Conclusion

This report presents the midline results of the mixed-methods cluster-RCT on the impact of IN's eSchool 360 model on ZAT, EGRA, EGMA, and Oral Vocabulary assessment scores 14 months after starting the program. To determine the impact of the program, we used a cluster-RCT in which 63 eligible schools were randomly assigned either to receive IN's eSchool 360 program (30 treatment schools) or not to receive the program (33 control schools). In addition, we conducted a process evaluation to determine the fidelity of program implementation.

The results of our study showed statistically significant ITT effects on EGRA, EGMA, ZAT, and Oral Vocabulary assessment scores. Specifically, we found statistically significant effects of 0.16 standard deviations or 3.1 percentage points on ZAT scores, 0.40 standard deviations or 3.5 percentage points on EGRA scores, 0.22 standard deviations or 4.9 percentage points on EGMA scores, and 0.25 standard deviations or 6.0 percentage points on Oral Vocabulary scores. The learning gains were equivalent to 1.88 additional years of education for reading and 1.03 additional years of education for mathematics if we extrapolate results from a sample of representative low- and middle-income countries (Evans & Yuan, 2019). The learning gains were equivalent to 1.2 additional years of education for reading and 0.37 additional years (or 4.5 additional months) of education for mathematics if we compare the ITT estimates in original units to the learning gains in the control group.

TOT effects showed larger impacts on ZAT, EGRA, EGMA, and Oral Vocabulary assessment scores for students who started attending the IN schools. Attending an IN school led to an increase of 0.32 standard deviation or 6.3 percentage points in the ZAT score, a 0.83 standard deviation or 7.2 percentage point increase in the primary EGRA score, a 0.45 standard deviation or 10.1 percentage point increase in the primary EGMA score, and a 0.52 standard deviation or 12.4 percentage point increase in the Oral Vocabulary score when we used attending IN schools 3 times in the week before the survey as a proxy for regular attendance in IN schools. Enrolling in an IN school led to improvements of 0.26 standard deviation or 5.2 percentage points in the ZAT score, a 0.68 standard deviation or 5.8 percentage points increase in the primary EGRA score, a 0.37 standard deviation or 8.2 percentage points increase in the primary EGMA score, and a 0.42 standard deviation or 10.1 percentage points increase in the Oral Vocabulary score when we used ever enrolling in IN schools as a proxy for enrollment.

Evidence suggested that increase in school enrollment and attendance, and improvements in the quality of education were likely the main drivers of the positive effects of the eSchool 360 model on learning outcomes. Children living in IN catchment areas were almost 8 percentage points more likely to be enrolled in schools compared to children in the control group, which translated into large increases in school attendance. In addition, among enrollees, caregivers in the IN catchment areas reported 0.30 points higher satisfaction with education (on a satisfaction scale), and 0.84 points higher level of engagement (on an engagement scale) in their child's education and school activities, compared with the control group. Parents' perceived quality of education and school attendance were also highly and statistically significantly correlated with EGRA, EGMA, ZAT, and Oral Vocabulary scores. However, we did not find statistically significant effects on child development outcomes.

In line with the quantitative research, all categories of qualitative respondents (students, teachers, parents, and program staff) showed a high perceived quality of education at IN schools. Many respondents alluded to their belief that IN schools provided a higher quality education than government schools and that teachers were present far more regularly at IN schools than at government schools. In agreement with our quantitative findings, qualitative respondents also perceived notable improvements in literacy as a result of the IN program, although they did not report similar improvements in numeracy.

Perhaps surprisingly, we also found positive effects of the eSchool 360 model on food security. We should exercise caution in interpreting this result, however, because we did not anticipate this mechanism. We will examine whether reductions in education expenditures (because of the free education offered by IN) enabled increases in food expenditures during the endline survey. The midline results already showed statistically significant reductions in education expenditures as a result of the eSchool 360 program, suggesting that the program may have enabled increases in food expenditures. We did not find statistically significant effects on parental-level aspirations, however. The eSchool 360 model did not show statistically significant effects on the probability parents assigned to their children relative to graduating from Grade 12, their children's age-at-marriage preferences, or the excepted bride price of their daughters.

A high level of fidelity of implementation likely contributed to the eSchool 360 model's positive effects on learning outcomes. Teachers used the participatory pedagogical approaches on which they were trained, adhered to the curriculum prescribed in the curriculum map, and used technology (tablets and projectors) as intended. While program implementation did not appear to vary based on geography or seasonality, we found evidence that student attendance may have been lower during rainy season and lower throughout the year for children who lived longer distances from school. We also saw evidence from the qualitative and quantitative data that females tended to attend school more regularly than males.

Although we found statistically significant effects on learning outcomes across the board, the results still suggested considerable scope for additional gains in learning outcomes. Despite the positive effects, children residing in IN catchment areas scored an average of only 10.9% on EGRA assessments and 24.3% on EGMA assessments. It will be critical to examine the longer-term effects of the eSchool 360 model to assess whether the program can achieve greater learning gains in future.

Further, it will be important to investigate the cost-effectiveness of the eSchool 360 model using a rigorous cluster-RCT. Previous nonexperimental research has shown that the eSchool 360 model may be cost-effective in achieving improvements in reading and mathematics outcomes. At present, however, it remains unclear whether a cluster-RCT will show the same results. A cluster-RCT of the outsourcing of education in Liberia showed that the program achieved large learning gains among some private providers, but program costs were considerable in some cases (Romero & Sandefur, 2019). Our cost-effectiveness analysis will examine whether the promise of the eSchool 360 model to achieve substantial learning gains for students in Zambian community schools for a moderate cost will hold true.

References

Alderman, H., Gilligan, D. O., & Lehrer, K. (2012). The impact of food for education programs on school participation in northern Uganda. *Economic Development and Cultural Change*, *61*(1), 187-218.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics, 122*(3), 1235–1264.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., … Wasserman, M. (2016, October). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* [NBER Working Paper No. 22746]. Cambridge, MA: National Bureau of Economic Research. Doi: 10.3386/w22746

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., … Udry, C. (2015). A multifaceted program causing lasting progress for the very poor: Evidence from six countries. *Science, 348*(6236). Doi: 10.1126/science.1260799

Berry, C., Barnett, E., & Hinton, R. (2015). What does learning for all mean for DFID's global education work? *International Journal of Educational Development, 40*, 323–329.

Central Statistics Office Zambia (2013). *2010 census of population and housing: Population and demographic projections 2011–2035.* Retrieved from http://www.zamstats.gov.zm/index.php/publications/category/11-2010-census

Central Statistical Office, Ministry of Health, & ICF International. (2014). *Zambia demographic and health survey: 2013–14*. Rockville, MD: Authors.

Chimese, K. (2014). *A case of vulnerables teaching OVCs.* Retrieved from http://zambiaopencommunityschools.org/news/a_case_of_vulnerables_teaching_ovcs_100011.php

Conn, K. M. (2014). *Identifying effective education interventions in sub-Saharan Africa: A meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Columbia University.) Retrieved from http://academiccommons.columbia.edu/catalog/ac%3A175014

De Hoop, T., Brudevold-Newman, A., & Davis, D. (2018). *Baseline report for the mixed-methods cluster-randomized controlled trial of Impact Network's eSchool 360 model in rural Zambia.* Washington, DC: American Institutes for Research.

De Hoop, T., Brudevold-Newman, A., Ring, H., Zanolini, A., Nowlin, C., & Davis, D. (2017). *Inception report for the mixed-methods cluster-randomized controlled trial of Impact Network's eSchool 360 model in rural Zambia.* Washington, DC: American Institutes for Research.

DeStefano, J. (2006). *Meeting EFA: Zambia community schools* [Educational Quality Improvement Program 2]. Washington, DC: United States Agency for International Development. Retrieved from https://www.epdc.org/sites/default/files/documents/Zambia%20Community%20Schools.pdf

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2011, August). *Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education*. Retrieved from http://economics.mit.edu/files/6959

Evans, D., & Yuan, F. (2019, February). *Equivalent years of schooling: A metric to communicate learning gains in concrete terms* [Policy Research Working Paper No. WPS 8752]. Washington, DC: The World Bank.

Filmer, D., & Pritchett, L. (2001). Estimating wealth effects without expenditure data—Or tears: An application to educational enrollments in states in India. *Demography, 38*(1), 115–132.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature, 40*(2), 436–482.

Heß, S. (2017). Randomization inference with Stata: A guide and software. The Stata Journal, 17(3), 630-651.

Imbens, G. (2015). Matching methods in practice. *Journal of Human Resources, 50*(2), 373–419.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science, 340*(6130), 297–300.

Ministry of General Education, Republic of Zambia. (2016, February). *Educational statistical bulletin*. Retrieved from https://www.moge.gov.zm/download/statistics/annual_statistics/Educational-Statistical-Bulletin/2014-Educational-Statistical-Bulletin.pdf

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, *109*(4), 1426–1460.

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Strigel, C. (2016). Does technology improve reading outcomes? Comparing the effectiveness and cost-effectiveness of ICT interventions for early grade reading in Kenya. *International Journal of Educational Development, 49,* 204–214.

Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Washington, DC: Center for Global Development.

Romero, M., & Sandefur, J. (2019). Beyond Short-term Learning Gains: The Impact of Outsourcing Schools in Liberia after Three Years. Center for Global Development Working Paper 521.

Schling, M., & Winters, P. (2018, July). Computer-assisted instruction for child development: Evidence from an educational programme in rural Zambia. *The Journal of Development Studies, 54*(7), 1121–1136.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., … Jimenez, E. (2016, September). *The impact of education programmes on learning and school participation in low- and middle-income countries: A systematic review summary report.* London, UK: International Initiative for Impact Evaluation. Retrieved from http://pubdocs.worldbank.org/en/429341481774674870/The-Impact-of-Education-Programmes-on-Learning-and-School-Participation.pdf

UNICEF. (2015). *The state of the world's children 2015: Executive summary.* Retrieved from http://www.unicef.org/publications/files/SOWC_2015_Summary_and_Tables.pdf.

United Nations. (2017). *The sustainable development goals report 2017*. New York, NY: United Nations Statistics Division. Retrieved from https://unstats.un.org/sdgs/files/report/2017/TheSustainableDevelopmentGoalsReport2017.pdf

United Nations Educational, Scientific, and Cultural Organisation. (2017). More than one-half of children and adolescents are not learning worldwide [Fact Sheet no. 46]. UNESCO Institute for Statistics. Retrieved from http://uis.unesco.org/sites/default/files/documents/fs46-more-than-half-children-notlearning-en-2017.pdf

Valk, J.-H., Rashid, A. T., & Elder, L. (2010, March). Using mobile phones to improve educational outcomes: An analysis of evidence from Asia. *The International Review of Research in Open* *and Distributed Learning, 11*(1), 117–140. Retrieved from https://files.eric.ed.gov/fulltext/EJ881581.pdf

Winters, C., Schling, M., & Winters, P. (2013). *Cost effectiveness report for the Impact Network eLearning education program in rural Zambia: Final report prepared for the Impact Network.* Unpublished manuscript.

World Bank. (2018). *World development report 2018: Learning to realize education's promise.* Washington, DC: Author. Retrieved from https://www.worldbank.org/en/publication/wdr2018

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. The Quarterly Journal of Economics, 134(2), 557–598.

Appendix A. Bivariate Correlations With School Enrollment

Table A-1. Demographic Characteristics and School Enrollment

	Not Enrolled at School		Enrolled at School		Difference Test			
Variables	Mean	*N*1	Mean	*N*2	Diff	SE	*p*-Value	Std. Mean Difference
Index child is female	0.30	715	0.57	985	0.27	0.03	0.00	0.54
Index child is 8 years old or older at baseline	0.51	715	0.50	985	-0.01	0.03	0.71	-0.02
Index child is a biological child	0.86	715	0.87	985	0.00	0.02	0.82	0.01
Household size	6.76	715	6.85	985	0.08	0.11	0.45	0.04
Caregiver has attended school	0.57	715	0.67	973	0.11	0.03	0.00	0.22
Highest level of education achieved by male household member (or caregiver when there is no male adult)	3.46	713	4.34	984	0.89	0.18	0.00	0.25
Katete district	0.13	715	0.18	985	0.05	0.03	0.17	0.12
Petauke district	0.58	715	0.56	985	-0.01	0.04	0.75	-0.03
Sinda district	0.29	715	0.26	985	-0.03	0.03	0.35	-0.07
Index child ethnicity: Chewa	0.37	709	0.38	978	0.01	0.04	0.75	0.03
Index child ethnicity: Nsenga	0.00	709	0.00	978	-0.00	0.00	0.44	-0.04
Index child ethnicity: Other	0.01	715	0.02	985	0.01	0.01	0.21	0.06
Index child language: Chewa	0.34	714	0.37	983	0.03	0.04	0.49	0.06
Index child language: Nsenga	0.65	714	0.63	983	-0.02	0.04	0.56	-0.05
Index child language: Other	0.00	715	0.00	985	-0.00	0.00	0.23	-0.07
Index child is the 1st child	0.14	666	0.15	807	0.01	0.01	0.28	0.03
Index child is the 2nd child	0.38	666	0.41	807	0.03	0.00	0.05	0.07
Index child is the 3rd child	0.35	666	0.32	807	-0.03	0.01	0.12	-0.07
Index child is the 4th child	0.11	666	0.11	807	-0.00	0.00	0.22	-0.00
Index child is the 5th child or older	0.03	666	0.01	807	-0.01	0.00	0.21	-0.08

Note. Standard errors (SE) clustered at the school level.

Table A-2. Household Socioeconomic Characteristics and School Enrollment

Variables	Not Enrolled at School		Enrolled at School		Difference Test			Std. Mean Difference
	Mean	*N*1	Mean	*N*2	Diff	SE	*p*-Value	
Household considers itself to be non-poor	0.01	715	0.02	985	0.00	0.00	0.34	0.04
Household considers itself to be moderately poor	0.45	715	0.54	985	0.09	0.03	0.00	0.17
Household considers itself to be very poor	0.54	715	0.44	985	-0.09	0.03	0.00	-0.18
Household food insecurity scale	14.12	715	12.75	985	-1.37	0.36	0.00	-0.21
Household asset index	-0.12	715	0.10	985	0.22	0.05	0.00	0.21

Note. Standard errors (SE) clustered at the school level.

Table A-3. Baseline Learning Outcomes and School Enrollment

Variables	Not Enrolled at School		Enrolled at School		Difference Test			Std. Mean Difference
	Mean	*N*1	Mean	*N*2	Diff	SE	*p*-Value	
Zambian Achievement Test (percentage correct)	0.41	715	0.47	985	0.06	0.01	0.00	0.25
Early Grade Reading Assessment (percentage correct)	0.04	715	0.05	985	0.01	0.00	0.00	0.20
Early Grade Math Assessment (percentage correct)	0.06	715	0.08	985	0.02	0.01	0.00	0.19
Oral Vocabulary (percentage correct)	0.58	715	0.62	985	0.04	0.01	0.00	0.16
Standardized Zambian Achievement Test (ZAT)	-0.12	715	0.08	985	0.20	0.05	0.00	0.20
Standardized Early Grade Reading Assessment (EGRA)	-0.15	715	0.11	985	0.25	0.05	0.00	0.25
Standardized Early Grade Math Assessment (EGMA)	-0.11	715	0.08	985	0.19	0.05	0.00	0.19

Standardized Oral Vocabulary	-0.09	715	0.07	985	0.16	0.05	0.00	0.16

Note. Standard errors (SE) clustered at the school level.

Appendix B. Adjusted *p*-Values for Multiple Comparisons in Heterogeneous ITT Effects

Table B-1. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Zambian Achievement Test (ZAT) (Standard Deviation)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	0.023	0.090	0.802	0.959
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.113	0.109	0.302	0.959
Treatment (ITT) X Caregiver attended school	-0.064	0.098	0.520	0.959
Treatment (ITT) X Index child scored above median ZAT at baseline	-0.015	0.114	0.896	0.959
Treatment (ITT) X Index child scored above median EGRA at baseline	0.007	0.129	0.959	0.959
Sample with children who scored above median in EGMA at baseline	0.027	0.118	0.818	0.959
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.159	0.103	0.126	0.959
Treatment (ITT) X Petauke	0.052	0.147	0.726	0.959
Treatment (ITT) X Sinda	0.051	0.174	0.772	0.959

Note. Standard errors clustered at the school level.

Table B-2. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Zambian Achievement Test (ZAT) (Percentage Correct)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	0.004	0.018	0.802	0.959
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.022	0.021	0.302	0.959
Treatment (ITT) X Caregiver attended school	-0.012	0.019	0.520	0.959
Treatment (ITT) X Index child scored above median ZAT at baseline	-0.003	0.022	0.896	0.959
Treatment (ITT) X Index child scored above median EGRA at baseline	0.001	0.025	0.959	0.959
Sample with children who scored above median in EGMA at baseline	0.005	0.023	0.818	0.959
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.031	0.020	0.126	0.959
Treatment (ITT) X Petauke	0.010	0.029	0.726	0.959
Treatment (ITT) X Sinda	0.010	0.034	0.772	0.959

Note. Standard errors clustered at the school level.

Table B-3. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Standardized Early Grade Reading Assessment (EGRA) (Standard Deviation)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	0.102	0.133	0.447	0.597
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.110	0.150	0.464	0.597
Treatment (ITT) X Caregiver attended school	0.070	0.134	0.604	0.604
Treatment (ITT) X Index child scored above median ZAT at baseline	0.132	0.105	0.214	0.432
Treatment (ITT) X Index child scored above median EGRA at baseline	0.246	0.153	0.113	0.340
Sample with children who scored above median in EGMA at baseline	0.338	0.131	0.012	0.063
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.283	0.112	0.014	0.063
Treatment (ITT) X Petauke	0.141	0.258	0.586	0.604
Treatment (ITT) X Sinda	0.295	0.248	0.240	0.432

Note. Standard errors clustered at the school level.

Table B-4. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Early Grade Reading Assessment (EGRA) (Percentage Correct)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	0.009	0.011	0.447	0.597
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.009	0.013	0.464	0.597
Treatment (ITT) X Caregiver attended school	0.006	0.012	0.604	0.604
Treatment (ITT) X Index child scored above median ZAT at baseline	0.011	0.009	0.214	0.432
Treatment (ITT) X Index child scored above median EGRA at baseline	0.021	0.013	0.113	0.340
Sample with children who scored above median in EGMA at baseline	0.029	0.011	0.012	0.063
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.024	0.010	0.014	0.063
Treatment (ITT) X Petauke	0.012	0.022	0.586	0.604
Treatment (ITT) X Sinda	0.025	0.021	0.240	0.432

Note. Standard errors clustered at the school level.

Table B-5. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Standardized Early Grade Mathematics Assessment (EGMA) (Standard Deviation)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	-0.004	0.107	0.972	0.972
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.068	0.106	0.524	0.817
Treatment (ITT) X Caregiver attended school	-0.059	0.123	0.633	0.817
Treatment (ITT) X Index child scored above median ZAT at baseline	0.103	0.100	0.305	0.817
Treatment (ITT) X Index child scored above median EGRA at baseline	0.051	0.114	0.659	0.817
Sample with children who scored above median in EGMA at baseline	0.061	0.122	0.621	0.817
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.136	0.105	0.200	0.817
Treatment (ITT) X Petauke	-0.216	0.197	0.277	0.817
Treatment (ITT) X Sinda	-0.069	0.197	0.726	0.817

Note. Standard errors clustered at the school level.

Table B-6. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Early Grade Mathematics Assessment (EGMA) (Percentage Correct)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	-0.001	0.024	0.972	0.972
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.015	0.024	0.524	0.817
Treatment (ITT) X Caregiver attended school	-0.013	0.028	0.633	0.817
Treatment (ITT) X Index child scored above median ZAT at baseline	0.023	0.022	0.305	0.817
Treatment (ITT) X Index child scored above median EGRA at baseline	0.011	0.026	0.659	0.817
Sample with children who scored above median in EGMA at baseline	0.014	0.027	0.621	0.817
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.030	0.023	0.200	0.817
Treatment (ITT) X Petauke	-0.048	0.044	0.277	0.817
Treatment (ITT) X Sinda	-0.016	0.044	0.726	0.817

Note. Standard errors clustered at the school level.

Table B-7. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Standardized Oral Vocabulary (Standard Deviation)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	-0.116	0.096	0.231	0.790
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.102	0.093	0.273	0.790
Treatment (ITT) X Caregiver attended school	-0.037	0.095	0.696	0.845
Treatment (ITT) X Index child scored above median ZAT at baseline	0.023	0.103	0.824	0.845
Treatment (ITT) X Index child scored above median EGRA at baseline	0.075	0.118	0.527	0.790
Sample with children who scored above median in EGMA at baseline	0.078	0.101	0.441	0.790
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.020	0.102	0.845	0.845
Treatment (ITT) X Petauke	-0.136	0.165	0.411	0.790
Treatment (ITT) X Sinda	-0.200	0.182	0.277	0.790

Note. Standard errors clustered at the school level.

Table B-8. Adjusted *p*-Values for Multiple Comparisons of Heterogenous Effects on Oral Vocabulary (Percentage Correct)

Sub-Groups	Coefficient	Standard Error	Unadjusted *p*-Value	Adjusted *q*-Value
Treatment (ITT) X Female	-0.028	0.023	0.231	0.790
Treatment (ITT) X Index child was 8 or 9 years old at baseline	0.025	0.022	0.273	0.790
Treatment (ITT) X Caregiver attended school	-0.009	0.023	0.696	0.845
Treatment (ITT) X Index child scored above median ZAT at baseline	0.006	0.025	0.824	0.845
Treatment (ITT) X Index child scored above median EGRA at baseline	0.018	0.028	0.527	0.790
Sample with children who scored above median in EGMA at baseline	0.019	0.024	0.441	0.790
Treatment (ITT) X Index child scored above median Oral Vocabulary at baseline	0.005	0.025	0.845	0.845
Treatment (ITT) X Petauke	-0.033	0.040	0.411	0.790
Treatment (ITT) X Sinda	-0.048	0.044	0.277	0.790

Note. Standard errors clustered at the school level.

Appendix C. Correlations between Intermediate and Learning Outcomes

Table C-1. Correlations Between Intermediate and Learning Outcomes for Complete Sample

	(1)	(2)	(3)	(4)
	ZAT (SMD)	EGRA (SMD)	EGMA (SMD)	Oral Vocabulary (SMD)
Index child was enrolled at any school (last year or this year)	0.376***	0.257*	0.368***	0.289**
	(0.089)	(0.101)	(0.079)	(0.091)
Number of days attended school last week	0.095***	0.154***	0.112***	0.046*
	(0.019)	(0.030)	(0.016)	(0.019)
Child development scale	0.061***	0.059***	0.065***	0.053***
	(0.014)	(0.010)	(0.009)	(0.012)
Petauke	-0.039	0.223	0.164	0.273*
	(0.084)	(0.138)	(0.104)	(0.120)
Sinda	0.082	0.269+	0.150	0.327**
	(0.078)	(0.142)	(0.092)	(0.111)
Highest level of education achieved by male household member	0.003	0.013	-0.008	0.002
	(0.007)	(0.012)	(0.007)	(0.007)
Index child age	0.093**	0.127**	0.147***	0.066*
	(0.028)	(0.041)	(0.027)	(0.027)
Index child is a female	0.024	0.056	0.019	0.069
	(0.051)	(0.065)	(0.045)	(0.057)
Index child is a biological child	-0.156*	-0.066	-0.192**	-0.224***
	(0.063)	(0.079)	(0.069)	(0.055)
Household size	0.015	-0.001	-0.007	0.006
	(0.013)	(0.018)	(0.014)	(0.015)
Caregiver has attended school	0.079	0.106+	0.103+	0.076
	(0.052)	(0.060)	(0.055)	(0.050)
Index child ethnicity: Tumbuka	0.443	-0.077	0.028	0.161
	(0.462)	(0.111)	(0.225)	(0.196)
Index child language: Chewa	0.074	-0.210	0.109	-0.274
	(0.193)	(0.563)	(0.318)	(0.293)

	(1)	(2)	(3)	(4)
	ZAT (SMD)	EGRA (SMD)	EGMA (SMD)	Oral Vocabulary (SMD)
Index child language: Nsenga	0.085	-0.396	0.103	-0.215
	(0.205)	(0.566)	(0.309)	(0.285)
Index child is the 1st child	0.138	0.160	0.076	0.150
	(0.147)	(0.173)	(0.151)	(0.138)
Index child is the 2nd child	0.098	0.175	0.057	0.078
	(0.143)	(0.166)	(0.152)	(0.121)
Index child is the 3rd child	0.077	0.153	0.097	0.093
	(0.142)	(0.150)	(0.152)	(0.134)
Index child is the 4th child	0.168	0.329	0.162	0.183
	(0.168)	(0.218)	(0.171)	(0.148)
Household food insecurity access scale	0.000	-0.012*	0.002	-0.008*
	(0.004)	(0.005)	(0.004)	(0.004)
Household asset index	0.023	0.046	0.040+	0.033
	(0.023)	(0.030)	(0.021)	(0.024)
Household distance from school (km)	-0.059+	0.022	-0.032	0.040
	(0.033)	(0.056)	(0.036)	(0.035)
Baseline standardized ZAT score	0.062+			
	(0.032)			
Baseline standardized EGRA score		0.043		
		(0.030)		
Baseline standardized EGMA score			0.125***	
			(0.028)	
Baseline standardized Oral Vocabulary score				0.079**
				(0.027)
Constant	-2.364***	-2.336***	-2.823***	-1.619***
	(0.404)	(0.635)	(0.444)	(0.446)
Observations	1659	1659	1659	1659
Adjusted R^2	0.220	0.170	0.266	0.143

Note. Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C-2. Correlations Between Intermediate and Learning Outcomes for the Sample Enrolled in School

	(1)	(2)	(3)	(4)
	ZAT (SMD)	EGRA (SMD)	EGMA (SMD)	Oral Vocabulary (SMD)
Number of days attended school last week	0.050*	0.125**	0.058*	0.016
	(0.023)	(0.045)	(0.024)	(0.022)
Child development scale	0.033*	0.062**	0.053**	0.022+
	(0.015)	(0.022)	(0.019)	(0.011)
Age at which index child first enrolled at school	-0.600***	-0.986***	-0.845***	-0.397***
	(0.065)	(0.103)	(0.055)	(0.064)
Caregiver perception/engagement scale	0.021	0.007	0.014	0.013
	(0.013)	(0.020)	(0.013)	(0.013)
Caregiver satisfaction scale	0.091**	0.279***	0.108**	0.059*
	(0.027)	(0.059)	(0.034)	(0.029)
Highest education level of male household member	-0.002	0.022	-0.005	0.008
	(0.010)	(0.017)	(0.009)	(0.007)
Petauke	-0.068	0.173	0.042	0.068
	(0.120)	(0.227)	(0.139)	(0.115)
Sinda	0.040	0.211	0.021	0.128
	(0.114)	(0.246)	(0.134)	(0.110)
Index child age	0.724***	1.184***	1.050***	0.484***
	(0.071)	(0.145)	(0.067)	(0.063)
Index child is a female	-0.025	0.024	-0.026	0.065
	(0.067)	(0.108)	(0.064)	(0.072)
Index child is a biological child	-0.235**	-0.092	-0.190+	-0.220***
	(0.082)	(0.123)	(0.102)	(0.059)
Household size	0.017	-0.038	-0.031	0.013
	(0.017)	(0.030)	(0.020)	(0.016)
Caregiver has attended school	0.087	0.148+	0.137+	0.084
	(0.058)	(0.083)	(0.075)	(0.065)
Index child ethnicity: Tumbuka	-0.144	-0.360	-0.381	0.143

	(1) ZAT (SMD)	(2) EGRA (SMD)	(3) EGMA (SMD)	(4) Oral Vocabulary (SMD)
	(0.313)	(0.609)	(0.295)	(0.203)
Index child language: Chewa	0.014	-0.989	-0.186	-0.383
	(0.323)	(1.145)	(0.630)	(0.477)
Index child language: Nsenga	-0.030	-1.261	-0.133	-0.278
	(0.323)	(1.146)	(0.623)	(0.479)
Index child is the 1st child	0.159	0.076	-0.121	0.121
	(0.218)	(0.369)	(0.307)	(0.196)
Index child is the 2nd child	0.098	0.168	-0.059	-0.028
	(0.197)	(0.346)	(0.288)	(0.186)
Index child is the 3rd child	0.130	0.067	-0.033	0.026
	(0.198)	(0.347)	(0.292)	(0.193)
Index child is the 4th child	0.196	0.673	0.146	0.220
	(0.232)	(0.440)	(0.324)	(0.192)
Household food insecurity access scale	-0.010+	-0.023**	-0.005	-0.016**
	(0.006)	(0.008)	(0.006)	(0.005)
Household asset index	-0.007	0.060	0.065+	0.012
	(0.027)	(0.056)	(0.038)	(0.032)
Baseline household distance from school (km)	-0.052	0.042	-0.046	0.048
	(0.047)	(0.102)	(0.060)	(0.036)
Baseline standardized ZAT score	0.058			
	(0.040)			
Baseline standardized EGRA score		0.033		
		(0.052)		
Baseline standardized EGMA score			0.156***	
			(0.037)	
Baseline standardized Oral Vocabulary score				0.083*
				(0.033)
Constant	-0.700	-0.481	-0.880	-0.129
	(0.572)	(1.229)	(0.743)	(0.659)

	(1) ZAT (SMD)	(2) EGRA (SMD)	(3) EGMA (SMD)	(4) Oral Vocabulary (SMD)
Observations	860	860	860	860
Adjusted R^2	0.168	0.171	0.242	0.123

Note. Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix D. ITT Effects on Probability of Non-Zero Scores

Table D-1. ITT Effects on Probability of Scoring Above Zero

	(1)	(2)	(3)	(4)
VARIABLES	ZAT	EGRA	EGMA	Oral Vocabulary
Treatment	0.006	0.047**	0.017**	0.022**
	(0.007)	(0.019)	(0.008)	(0.009)
Observations	1,688	1,688	1,688	1,688
R-squared	0.006	0.016	0.012	0.011
Control group mean	0.981	0.834	0.963	0.967

Note. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH

1000 Thomas Jefferson Street NW

Washington, DC 20007-3835 | 202.403.5000

www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC Austin, TX | Arlington and Reston, VA | Seattle, WA

International: Algeria | Ethiopia | Germany | Haiti | Zambia

8908_02/20